University of Alberta

NMR Structure and Functional Studies of two proteins from Methanobacterium thermoautrophicum

by

©

Godwin Yao Amegbey

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in

Pharmaceutical Science

Faculty of Pharmacy and Pharmaceutical Sciences

Edmonton, Alberta

Fall 2005

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

0-494-08606-8

Your file Votre référence
ISBN
Our file Notre reference
ISBN

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

This project was under taken as part of the 'Structural Proteomics Initiative' from the Ontario Cancer Institute. The Structures of two proteins (MTH0807 and MTH0776) from *Methanobacterium thermoautotrophicum* (strain ΔH) was determined by NMR spectroscopy to aid the prediction of their function. Using structure analysis and visualization tools in addition to modern NMR spectroscopy and comparative biochemical assays, we have shown (in chapter 2) that it may be possible to engineer multiple functions (thioredoxin and glutaredoxin activities) into MTH0807. This will be particularly useful in the technological application of MTH0807 as it is very stable and seems to withstand the deleterious effects that most proteins could not tolerate. In chapter 3, our analysis also indicated that MTH0807 is indeed a true thioredoxin, contrary to earlier suggestions by previous workers that MTH0807 may be part of a distinct ribonucleotide-reducing system involving ferredoxins and ferredoxin-thioredoxin reductases. In chapter 4 of this thesis, the three dimensional structure of MTH0776, a 101 amino acid residue protein specific to the methanogenic archaebacteria with no known function or structure was determined. It exhibited a novel fold hence no known function could be predicted from its structure. Our extensive comparative genomic analysis however shows that MTH0776 is part of a two-protein operon with its upstream partner -- MTH0777, which is also specific to methanogenic archaebacteria. The fact that this two-component operon is conserved in all methane metabolizing archaea, strongly suggests that both proteins have an important role in methanogenesis. Analysis

shows that MTH0776 and MTH0777 bind together and may be part of a transport system conserved in the methanogenic archaebacteria.

Acknowledgements

I would like to express my gratitude to my supervisor, Dr David S. Wishart, for his encouragement, support, guidance and constructive criticism. Without his encouragement throughout this program, I would never have been able to accomplish this work.

I wish to express my whole-hearted thanks to all my colleagues and fellow students to whom I have had the pleasure of associating with for the past six years. In particular, I would like to thank Dr Hassan Monzavi who has been a tremendous friend, providing scientific assistance with regard to NMR issues, many intellectual conversations and a continual source of moral support. I would like to thank Rajarshi Maiti, Haiyang Zhang and Nelson Young, whose help in regards to computer related issues has been tremendous. I would also like to thank Trent Bjorndahl, Lena Andrew and Ashenafi Abera for their useful discussions in relation to my work.

I am grateful to my supervisory committee members, Professors Mavanur Suresh, John Samuel, Steve McQuarrie and George Kotovych for their advice which improved the content of this thesis. As well, the financial support from the Canadian Commonwealth Scholarship Scheme, Pfizer Graduate Student Scholarship and PENCE is gratefully acknowledged.

Finally, I am extremely grateful to my wife Damaris, my daughter Josephine and son Andrew for their endless patience, love, sacrifice and moral support for the past six years.

Table of Contents

Page No

Chapter 1: Introduction 1

1.1 Structural Proteomics and drug discovery........................ 1

1.2 Structural Proteomics of an archaeon............................... 4

1.3 NMR methodology in the structure generation process 7

1.4 Pharmaceutical potential of Thioredoxins and Glutaredoxins... 21

1.5 Thesis outline and description of work done..................... 24

1.6 References.. 26

Chapter 2: Assay and insights into T4 glutaredoxin dual activity in comparison to other thioltransferases 43

2.1 Introduction 43

2.2 Materials and methods... 45

2.3 Results.. 46

2.4 Discussion 47

2.5 Conclusion 56

2.6 References.. 57

Chapter 3: Structural and functional characterization of MTH0807. 65

3.1 Introduction.. 65

3.2 Materials and methods.. 67

3.3 Results 72

3.4 Discussion 79

3.5 Conclusion.. 87

3.6 References 88

Chapter 4: Structural and functional characterization of MTH0776.. 104

4.1 Introduction 104

4.2 Materials and methods 105

4.3 Results 111

4.4 Discussion and conclusion 114

4.5 References.. 121

Chapter 5: Conclusion and future studies 144

5.1 Conclusion and future studies.. 144

5.2 References 150

Appendix 1: Statistical analysis of thioredoxin/glutaredoxin activity.. 151

Appendix 2: Measurement of T_2 Relaxation 155

List of Tables

Table — Page No

1.1 Eukaryotic and Prokaryotic traits found in Archaebacteria 35

1.2 Function of thioredoxin in different organisms 36

2.1 Structure comparison of T4 Glutaredoxin and other oxidoreduxtases 59

2.2 Solvent accessible surface area of unique hydrophobic groove 60

3.1 Structural statistics for MTH0807 95

3.2 Structures similar to MTH0807 derived from SCOP database 96

4.1 Structure statistics for ensemble of 20 MTH0776 structures 126

1A1 Single sample t-test analysis of glutaredoxin activity 152

1A2 Single sample t-test analysis of thioredoxin activity (Oxidation of NADPH) 153

1A3 Single sample t-test analysis of thioredoxin activity (Reduction of DTNB) 154

2A1 Table of ^{15}N T_2 relaxation times of MTH0776 155

2A2 Table of ^{15}N T_2 relaxation times of MTH0776 in complex with MTH0777 156

2A3 Table of assigned chemical shifts of MTH0807 157

2A4 Table of assigned chemical shifts of MTH0776 160

List of Figures

Figure Page No

1.01 Precession of spin angular momentum of a nucleus with non-zero spin 37

1.02 Energy level diagram for a proton with spin=½ in a magnetic field 37

1.03 Net magnetization generated by population difference in applied field 38

1.04 Application of a 90° pulse along the x-axis with respect to Bo 38

1.05 Fading of the XY magnetization (relaxation) after a 90° pulse 39

1.06 Conversion of FID to the frequency domain by Fourier transformation 40

1.07 Detected Nuclei in an HNCA tripple resonance experiment 40

1.08 Detected Nuclei in an HNCACB tripple resonance experiment 40

1.09 Detected Nuclei in an HCCH-TOCSY tripple resonance experiment 41

1.10 Detected Nuclei in an HNCO tripple resonance experiment 41

1.11 Detected Nuclei in an HNHA 3D heteronuclei experiment 41

1.12 Detected Nuclei in an C(CO)NH tripple resonance experiment 42

1.13 Detected Nuclei in an H(CCO)NH tripple resonance experiment 42

2(a) Measurement of Glutaredoxin activity kinetics by UV absorbance 61

2(b) Measurement of Thioredoxin activity kinetics (oxidation of NADPH) 61

2(c) Measurement of Thioredoxin activity kinetics (reduction of DTNB) 62

2.1 Backbone superposition of the four stranded β-sheet in Thioredoxins 63

2.2 Backbone superposition of the active site C-X-X-C in Thioredoxins 63

2.3 Structure-based sequence alignment of T4 Grx and other oxidoreductases 64

3.1 pET-15b cloning/expression system from Novagen for MTH0807 97

3.2 ^{1}H-^{15}N HSQC spectrum for MTH0807 98

3.3 Backbone atom superposition of 20 structures of MTH0807 99

3.4 Ribbon diagram of a representative structure of MTH0807 100

3.5 Sequence alignment of MTH0807 and other oxidoreductases 101

3.6 Measurement of thiol ionization in MTH0807 by UV absornace 102

3.7(a) Measurement of Glutaredoxin activity kinetics by UV absorbance 102

3.7(b) Measurement of Thioredoxin activity kinetics (oxidation of NADPH) 103

3.7(c) Measurement of Thioredoxin activity kinetics (reduction of DTNB) 103

4.01 pET-15b cloning/expression system from Novagen for MTH0776 127

4.02 ^{1}H-^{15}N HSQC spectrum for mixture of MTH0776 and MTH0777 128

4.03 ^{1}H-^{15}N HSQC spectrum for MTH0776 129

4.04 A BLAST search for protein related to MTH0776 130

4.05 ClustalX sequence alignment of proteins related to MTH0776 130

4.06 Genome map of *Methanobacterium thermoautotrophicum* (ΔH) 131

4.07 Genome map of *Methanosarcina mazei* (Goe1) 132

4.08 Genome map of *Methanococcus jannaschii* 133

4.09 Genome map of *Methanococcus maripaludis* (S2) 134

4.10 Genome map of *Methanopyrus kandleri* (AV19) 135

4.11 Genome map of *Methanosarcina acetovorans* (C2A) 136

4.12 Phylogenetic tree of sequences related to MTH0776 137

4.13 Superimposed ^{1}H-^{15}N HSQC spectrum for pure MTH0776 and MTH0776/MTH0777 mixture 138

4.14 CPK model of MTH0776 showing perturbed residues in protein mixture 139

4.15 Ribbon diagram of a representative structure of MTH0776 140

4.16 Backbone atom superposition of 20 structures of MTH0776 141

4.17 Residues forming the two largest pockets on the surface of MTH0776 142

4.18 Surface rendering of the two pockets on MTH0776 143

2A1a Histogram distribution of ^{15}N T_2 Relaxation times of MTH0776 164

2A1b Graphs of ^{15}N T_2 Relaxation times of MTH0776/MTH0777 complex 165

2A2 Summary of observed short range NOEs, CSI and sequence number 166

2A3 Comparison of ^{1}H-^{15}N HSQC spectra in T_2 measurements of MTH0776 and MTH0776 in complex with MTH0777 167

List of Abbreviations

Amino acids:

A (Ala) – L-Alanine

C (Cys) – L-Cysteine

D (Asp) – L-Aspartic acid

E (Glu) – L-Glutamic acid

F (Phe) – L-Phenylalanine

G (Gly) – L-Glycine

H (His) – L-Histidine

I (Ile) – L-Isoleucine

K (Lys) – L-Lysine

L (Leu) – L-Leucine

M (Met) – L-Methionine

N (Asn) – L-Asparagine

P (Pro) – L-Proline

Q (Gln) – L-Glutamine

R (Arg) – L-Arginine

S (Ser) – L-Serine

T (Thr) – L-Threonine

V (Val) – L-Valine

W (Trp) – L-Tryptophan

Y (Tyr) – L-Tyrosine

2D – Two Dimensional

3D – Three Dimensional

BMRB – BioMagResBank

CSI – Chemical Shift Index

DNA – Deoxyribonucleic Acid

DQF-COSY – Double Quantum Filtered Correlation Spectroscopy

DSB – Disulfide Bond Forming Protein

DSS – 2, 2 -dimethyl-2-silapentane-5-sulfonic acid

DTNB – 5, 5' –dithiobis (2-nitrobenzoic acid)

DTT – Dithiothreitol

E. coli – Escherichia coli

EDTA – Ethylenediamine – tetra acetic acid

FID – Free Induction Decay

FT – Fourier Transformation

Grx – Glutaredoxin

GSH – Reduced Glutathione

GSSH – Oxidised Glutathione

HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HSQC – Heteronuclear Single Quantum Correlation

IPTG – Isopropyl-β-D-thiogalactopyranoside

kDa – Kilodalton

MTH – Methanobacterium thermoautotrophicum (strain ΔH)

NADP – Nicotinamide Adenine Dinucleotide Phosphate (oxidised)

NADPH – Nicotinamide Adenine Dinucleotide Phosphate (reduced)

NOE – Nuclear Overhauser Enhancement

NOESY – Nuclear Overhauser Enhancement Spectroscopy

NMR – Nuclear Magnetic Resonance

OD_{600} – Optical Density at 600 nm

PDB – Protein Data Bank

PDI – Protein Disulfide Isomerase

RDC – Residual Dipolar Coupling

RF – Radio Frequency

RMSD – Root Mean Square Deviation

RNA – Ribonucleic Acid

ROS – Reactive Oxygen Species

SAR – Structure Activity Relationships

TNB – 5-thio (2-nitrobenzoic acid)

TOCSY – Total Correlation Spectroscopy

TROSY – Transverse Relaxation Optimized Spectroscopy

TF – Transcription Factor

Trx – Thioredoxin

UV – Ultraviolet

Chapter 1

Introduction

1.1 Structural Proteomics and Drug Discovery

Over the past 10 years there has been a growing flood of new protein sequence information due to the completion or near completion of literally hundreds of genome projects (*1*). These sequencing efforts are producing complete descriptions of the protein repertoires in organisms that extend from bacteria to humans (*1*). For this sequence information to be truly useful, biologist are now faced with the daunting task of determining the molecular and cellular functions of hundreds of thousands of 'new' proteins which do not appear to have detectable sequence homology to any protein with known function. Given that proteins are the molecular engines of living cells, the importance of knowing the precise role a given protein plays in the cell can not be over emphasized. This information is important not only for a basic understanding of biology, but also for the discovery of both therapeutic targets and therapeutic proteins. For instance, the sequencing of the human, mouse and rat genomes has led to the identification of more than 10,000 potential therapeutic targets (*2*), while the sequencing of more than 200 bacteria and 2000 viruses has led to hundreds of potential drug targets. However, a gap has been created between genomics and drug discovery reflecting the fact that in less than 50% of sequenced genes, the gene sequence reveals little about the protein function or disease relevance (*2, 3*).

The prevailing method of predicting a 'new' protein's function in the absence of experimental data is the pairwise comparison of its sequence to previously sequenced proteins of known function. The more similar the sequence (> 30%), the more similar the function is likely to be (*4*). The pair of proteins are said to be orthologous if the sequence similarity is high (> 30%), they perform the same function but are present in different organisms. They are described as paralogous if the sequence similarity is high; they perform the same function but are present in the same organisms. This method has a major disadvantage in that structure and function (the two go hand –in – hand for proteins) are much more conserved than sequence (*5*). Proteins sharing similar functions often have similar folds, a result of descent from a common ancestral protein (*4*). Fold similarity can be measured in terms of RMSD, Z-Scores, P-Scores or Q-Scores (*6*). Z-score measures the statistical significance of a match in terms of Gaussian statistics. The higher the Z-score (>5), the higher is the statistical significance of the match. RMSD is the Root Mean Square Deviation, calculated between Cα-atoms of matched residues at best 3D superposition of the query and target structures. RMSD is presented in angstroms (Å). In simple words, RMSD gives you an idea of how separated, at best 3D superposition, a "typical" pair of matched Cα-atoms is. Generally, the larger the RMSD (>3.5), the more distant the matched structures are. P-score represents the minus logarithm of the P-value. P-value measures the probability of achieving the same or better quality of match at a chance, i.e. at random picking the structures from the database. It takes into account RMSD, number of aligned residues, number of gaps and number of matched Secondary Structure Elements. The higher the P-score (>3), the more surprising, or statistically significant, is the match. It should however be noted that finding a fold match between a target protein and one in the database does not always provide a reliable prediction of a protein's function as a result of functional evolution in which the same fold is found to support several different functions. An example is the

TIM-barrel fold which supports over 60 different functions (*4*). The fold match however usually suggests a possible function type and it's the first stop for structure – based function prediction. In this approach one tries to predict the molecular function of a protein on the basis of its three dimensional structure. While the number of different protein sequences may number in the billions, most biologists agree there are probably less than two thousand distinctly different types of protein folds in nature and that the majority of all known (and unknown) sequences will belong to one of these folds (*7*). Given that the number of folds is much less than the number of sequences, and given that the number of folds is comparatively small, some structural biologists have suggested that a major effort should be made to determine the 3D structures of all 2,000 fold "families". This has led to the concept of 'structural proteomics' or 'structural genomics', the determination of three dimensional protein structures on a genome-wide scale (*8*).

Fundamentally, structural proteomics is aimed at determining the sufficient number of three dimensional protein structures necessary to define a 'basic parts' list of all protein folds (*9, 10*). From this list most other protein structures could be modelled using computational techniques such as homology modelling (*9, 11*). Most protein chemists believe this three dimensional structural information can then be used to uncover clues to protein function that are not detectable from simple sequence comparisons (*12, 13*). Structural proteomics has the potential of appreciably increasing the number of protein drug targets as well as impacting heavily on the design of new pharmaceuticals (*14*). For the pharmaceutical industry, structural information can be used to ascribe function, thereby revealing potentially new drug targets. Structural information can also be used to validate targets based on homology to other proteins known to bind specific small molecules. Structural information can also be used to invalidate targets by identifying key structural properties that do not lend themselves to drug binding. It is expected that as more protein structures become available and more of their functions become known, there will be an increase in the rate at which lead molecules for modulating target function are produced and optimised, ultimately generating an increased flow of drug candidates to the clinics.

1.2 Structural Proteomics of an Archaeon

In 1999, a prototype structural proteomics study of the proteins from the proteome of *Methanobacterium thermoautotrophicum* (strain ΔH) was organised by the Ontario Cancer Institute and the Department of Medical Biophysics, University of Toronto (*15*). The basic idea was to create a central production/screening facility for this archaeon's soluble proteins and then to 'farm out' interesting or well behaved proteins (of unknown function) to X-ray crystallographers and NMR spectroscopist across Canada and the United States. The primary goals of this research program were to evaluate the technical hurdles involved in such a high-throughput project and to determine if a distributed structure determination approach could work. Additionally the consortium wanted to estimate the percentage of proteins encoded by the genome that are immediately amenable to structure analysis and to assess the extent to which function can be inferred from structure (*15*). A retrospective analysis done by the researchers on the experimental behaviour of these proteins suggested some simple relations between sequence and solubility (*16*). This strongly suggested that databases of protein properties could be particularly useful in optimizing high throughput strategies such as NMR structure studies. An added bonus was that many of the 3D structures determined initially provided strong clues about the biochemical functions -- functions that were not detectable from sequence analysis. In many cases these putative functions suggested by the 3D structure could be readily confirmed by simple biochemical assays (*15, 16*). These encouraging results demonstrated that structural proteomics is feasible and can play a central role in functional genomics. Work has continued on this project ever since.

Methanobacterium thermoautotrophicum (strain ΔH) is an archaeabacterium. Archaea are a distinct branch or Kingdom of life, different from eukaryotes and prokaryotes. The distinction between prokaryotes and eukaryotes are that eukaryotic organisms share such characteristics as nuclei, cytoskeleton and internal membranes while prokaryotes do not (*17*). The archaeabacteria are microscopic prokaryotes but have a curious mix of traits, characteristic of both prokaryotes and eukaryotes as summarised in Table 1.1 (*17, 18*). Archaeabacteria have increasingly become the study of scientific investigation as in many ways, archaeal cells resembles the cells of bacteria but in a number of important respects they are more like cells of eukaryotes (*18*). Comparative studies on sequences of duplicated genes by scientists have found that archaeabacteria may actually be more closely related to eukaryotes than eubacteria (*18-21*). Archaeans include inhabitants of some of the most extreme environments such as rift vents in the deep sea at temperatures well over 100 °C, hot springs or in extremely alkaline or acidic waters. They have been found thriving inside the digestive tracts of cows, termites and marine life where they produce methane. They live in the anoxic muds of marshes and some can survive the desiccating effects of extremely saline waters (*22, 23*). The

Archaeans are normally considered by biologist to lie near the "root" of the tree of life because of their ability to live in extreme environments as well as their ability to make food using primordial materials (H_2, S, CO_2) from the earths' crust (*18*). Because some archaea have enzymes that function at high temperatures, considerable effort is being made to exploit them for commercial processes such as providing enzymes to be added to detergents (to maintain activity at high temperatures and pH) as well as being enlisted to aid in cleaning up contaminated sites example petroleum spills (*19*). Taq polymerase, a DNA polymerase isolated from *Thermus aquaticus* (which lives in hot springs) is used in the Polymerase Chain Reaction (PCR) for DNA replication/multiplication.

M. thermoautotrophicum was originally isolated in 1971 from a sewage pond in Urbana, Illinois (*24*). It is lithoautotrophic (i.e. it produces energy by modifying inorganic compounds such as minerals and assimilates carbon dioxide as its sole source of carbon) (*25*), thermophilic, and grows at an optimal temperature of 65°C. It biodegrades waste materials and produces natural gas (methane) in the process. Due to its potential economic and environmental importance the *M. thermoautotrophicum* genome was sequenced in the mid 1990's (*26)* to better understand the underlying biochemistry and microbiology of its unique biodegradation capabilities. The genome (1,751,377 bp) consists of 1855 open reading frames, with roughly 2/3 of the proteins having assigned functions. Because of the favorable solution properties of proteins isolated from thermophilic archaebacteria, *M. thermoautotrophicum* has become one of the most actively studied organisms in structural genomics (*16*). Furthermore, since *M. thermoautotrophicum* is a primitive archaeon near the 'ground floor' of evolution, the information obtained on its protein folds could be transferable to the proteins of many "higher" organisms (including eukaryotes).

My supervisor's laboratory is one of the many NMR laboratories involved in this project and it is our hope that our efforts would shed some light into the unknown gene products of this organism, ultimately producing structures for targets shown to be essential for bacterial growth and survival. We have so far completed work on three proteins (MTH0895, MTH0807 and MTH0776) of which two will be discussed in this thesis.

1.3 NMR Methodology in the Structure Generation Process

To date, the most prominent technique for 3D structure determination has been X-ray crystallography (*27*). No other method of structure determination can match X-ray crystallography's capacity for generating the structures of important proteins or protein complexes such as polymerases, proteosomes, viruses, and ribosomes. However, not all proteins are conducive to crystallization and not all proteins behave or look the same in a crystalline state as they do in the cellular milieu (*28*). In this regard, nuclear magnetic resonance (NMR) spectroscopy has emerged as an important alternative to X-ray crystallography as it allows protein structures to be determined in conditions that are very close to the physiological state (i.e. in solution). What's more, NMR spectroscopy provides structural biologists the opportunity to measure events or processes that cannot readily be seen or quantified by X-ray crystallography, such as protein kinetics, dynamics or thermodynamics (*28-30*).

Although size limitation (< 25kDa) and water solubility properties of proteins are disadvantages in 'conventional' NMR structure determination, recent advances have made such problems trivial. Recent advances in the NMR spectroscopy techniques such as partial deuteration of protein samples, stronger magnets (up to 900MHz), cold probes, solid state NMR, TROSY experiments (*31*), triple resonance methods for resonance assignments (*32, 33*), reverse micelles for handling membrane proteins, residual dipolar coupling analysis (*34*) along with automated and semi-automated methods for resonance assignments and structure determination (*35, 36*) have greatly increased the speed and accuracy of structure determination. These technical advances are proving to be critical to the success of the high-throughput structural proteomics projects (*10*).

NMR is also proving to be an indispensable tool for functional assignment and target validation due to its unique ability to rapidly and accurately obtain information about macromolecular dynamics and ligand binding properties. For instance, the technique known as Structure Activity Relationships (SAR) by NMR has recently been used to screen a library of 15,000 compounds, with minimum 1 mM sensitivity in the determination of the dissociation constants (*37*). With recent improvements in solid state NMR and membrane protein handling, NMR may eventually contribute significantly to structural proteomics of membrane proteins (*10*). However, the requirement of abundant quantities of isotopically labeled proteins and inherently poor expression levels of many membrane proteins may still be significant rate-limiting factors.

1.3.1 Theoretical Principles of NMR

NMR spectroscopy deals with the interaction between the magnetic moments of atomic nuclei and a magnetic field (*38*). The magnetic moment of a nucleus (μ) is intimately connected to an intrinsic property called spin angular momentum $\boldsymbol{I}$, whose magnitude is quantized. According to quantum mechanics each subatomic particle (e.g., proton, neutron, electron), has a spin value of ½. Therefore, the overall spin property of an atom or nucleus results from a combination of the spins of its constituent subatomic particles. Some common nuclei, notably ^{12}C and ^{16}O, with even number of protons and neutrons, have $I = 0$, (i.e. no spin angular momentum), no magnetic moment and consequently are not NMR active. Isotopes with an odd number of neutrons and an even number of protons or vice-versa (e.g., ^{1}H, ^{15}N, etc) usually have a half-integral quantum number and generally produce excellent NMR spectra. Nuclei with odd numbers of protons and neutrons (e.g., ^{14}N) have more complex spin states and are less suitable for direct NMR observation. Fortunately, each of the four most abundant elements in biological molecules (H, C, N, and O) have one naturally occurring isotope with a non-zero nuclear spin and, therefore, are observable in an NMR experiment.

While the naturally occurring isotope of hydrogen, ^{1}H is present at >99% abundance, other NMR active nuclei like ^{13}C and ^{15}N are present at much lower abundance (1.1% and 0.4%, respectively) and can be observed only after the target molecules are isotopically enriched. The ^{17}O isotope of oxygen, however, does not produce good NMR spectra as with a spin number $I = 5/2$, it has a non-spherical nuclear charge distribution giving rise to a quadrupole moment. This affects the relaxation time and consequently, significantly broadens the line width of the signal.

In the presence of an external magnetic field, the spin angular momentum of a nucleus with non-zero spin will cause that nucleus to undergo a cone-shaped rotational motion called 'precession' (Figure 1.1). The rate of precession, referred to as Larmour frequency (ω_0) is dependent on the strength of the external magnetic field (B_0) and intrinsic properties of the nucleus reflected in its gyromagnetic ratio (γ).

$$\omega_0 = - \gamma B_0 \tag{1.1}$$

Each magnetic nucleus has $2I+1$ possible spin orientations and $2I+1$ corresponding energy levels with respect to an external magnetic field. For example, a spin - ½ nucleus (^1H, ^{13}C) has two possible orientations, parallel and antiparallel, which corresponds to two different energy levels (Figure 1.2). The energy difference between these two levels is directly proportional to the strength of the magnetic field.

$$\Delta E = \gamma h B_0 / 2\pi \tag{1.2}$$

where h is Planck's constant. Note that in the absence of a magnetic field the spin angular momentum has no preferred directions and therefore leads to no detectable energy difference.

When placed in a magnetic field, a collection of magnetic nuclei, each absorbing a discrete amount of energy at its Larmour frequency, will partition themselves amongst the $2I+1$ available energy levels according to the Boltzmann distribution.

$$N_\beta / N_\alpha = e^{-\Delta E / kT} \tag{1.3}$$

where N_β and N_α are the population of the lower and upper states respectively.

This population difference generates a net magnetization (M), which aligns with the external magnetic field (B_0) and remains in this equilibrium state (Figure 1.3). If a magnetic pulse is applied for a short period of time in such a way that it produces a second magnetic field (B_1) perpendicular to the static field (B_0), it will drive M away from its equilibrium position by a so-called "flip angle". The magnitude of the flip angle depends on the time period and the field strength of B_1. A 90° pulse is therefore the time it takes for particular field strength to rotate the equilibrium magnetization 90° with respect to its equilibrium direction (which is the applied magnetic field usually set along the Z axis – see Figure 1.4). Placing a metal coil in the XY-plane allows the recording of the oscillating current generated by the precessing magnetization. The precessing magnetization eventually falls back to equilibrium, with the XY magnetization slowly fading and the Z magnetization growing (Figure 1.5). This oscillating magnetic field can be detected by an RF (radio frequency) coil, converted to an electric signal and recorded. This signal is called the free-induction decay (FID). The free induction decay can be converted from the FID (time-domain) to the frequency domain via Fourier transformation (Figure 1.6).

1.3.2 Chemical Shifts

The magnetic field at a nucleus is not exactly equal to the applied magnetic field. The electrons in a molecule surrounding a nucleus create a small magnetic field, which shields the nuclei slightly from the external field. The degree of shielding depends on whether a neighbouring chemical group pushes or withdraws electron density from the nucleus through various inductive effects. Therefore, the Larmor frequencies of different nuclei vary due to their different chemical environment. This frequency change is called the chemical shift and it is an exquisitely sensitive indicator of chemical structure and composition. The chemical shift (δ) is typically presented or measured in ppm (parts per million) and is defined as:

$$\delta = \frac{\text{Shift from standard (Hz)} \times 10^6}{\text{Spectrometer frequency (Hz)}} \text{ ppm} \ . \qquad (1.4)$$

NMR spectroscopists prefer to measure chemical shifts in ppm instead of Hz because the former is independent of the magnetic field strength B_0. The chemical shift reference most commonly used is the signal of the methyl groups of tetramethylsilane (TMS), which, by definition resonates at 0 ppm. In protein NMR 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) is used equivalently. When referencing to ^{13}C, ^{15}N and even ^{31}P, most NMR spectroscopists use an indirect referencing method based on the DSS signal (*39*). Different chemical groups have different chemical shifts and consequently chemical shift assignments provide a great deal of information for NMR spectroscopists. In proteins, for example, the signals of HN, Hα, aromatic and aliphatic protons, all the backbone and side chain ^{13}C atoms and also ^{15}N signals in isotopically labelled proteins can easily be distinguished and frequently assigned on the basis of their chemical shifts. Additionally, the chemical shifts contain valuable information about protein secondary structure (*40*).

1.3.3 Scalar Coupling

Nuclei, when close to one another, exert an effect on each other's effective magnetic field. This effect manifests itself in the NMR spectrum when the nuclei are experiencing different chemical environments or are chemically non-equivalent. If the separation between non-equivalent nuclei is less than or equal to three (sometimes four) bond lengths, this effect is observable as a form of peak splitting or through the appearence of 'multiplets'. The separation between the lines in a multiplet (in Hz) is given by the coupling constant – J, which is independent of the magnetic field strength. J couplings between pairs of protons separated by three covalent bonds (designated as ^3J or vicinal coupling), contains information about the intervening torsion angles and is described by the Karplus equation (*41*):

$$^3J = A\cos(\theta) + B\cos^2(\theta) + C \qquad (1.5)$$

where A, B and C are empirically derived constants and θ is the torsion (dihedral) angle. For proton NMR the three-bond coupling constant between the intra-residue alpha and amide protons ($^3J_{HNH\alpha}$) is most useful for protein secondary and tertiary structure determination as it can directly be related to the backbone dihedral angle ϕ (*42*).

1.3.4 Protein Structure Determination by NMR

Protein structure determination is a step-wise process and can be divided into four steps: 1) sample preparation; 2) data collection; 3) sequential assignment and spectral validation; and finally 4) structure generation and validation. Protein samples may be either unlabeled (^1H only) or labeled with certain NMR sensitive isotopes (i.e. ^{13}C and ^{15}N). Because ^1H is a naturally occurring and highly sensitive NMR nucleus, many smaller peptides and proteins need not be isotopically labeled. In fact, for the first 10 years of protein NMR, most proteins were not labeled (*42*). It has only been with the advent of better NMR hardware and NMR pulse sequences that the trend towards isotopically labeled samples has taken off (*28, 43, 44*). For larger proteins (>10 kD) or in the pursuit of higher quality structures, most NMR spectroscopists like to work with isotopically labeled proteins. These are prepared by growing cells and expressing the protein in defined minimal media that has been enriched with ^{13}C and/or ^{15}N labeled substrates.

Once a protein sample is ready for NMR (labeled or unlabeled) and suitable solution conditions found, it is then subjected to a series of 2D and/or 3D NMR experiments. The intent of these experiments is to collect sufficient spectral data so that every NMR peak (or resonance) can be assigned to every NMR detectable atom (or nucleus) in the protein (*29, 42*). This process is called sequential assignment. For unlabeled peptides and proteins, the assignment process is relatively straightforward. Two types of 2D NMR experiments are collected. A TOCSY (total correlation spectroscopy) experiment is first collected which allows individual amino acids to be identified by the distinct patterns of their chemical shifts (*45*). A NOESY (nuclear overhauser enhancement spectroscopy) experiment is then collected which yields a spectrum that looks almost identical to the TOCSY spectrum (*46*). However, in the NOESY experiment there are extra peaks (often 100's, even 1000's). These extra "pseudo" peaks or cross peaks, which show up between pairs of TOCSY peaks, correspond to interactions where nearby protons are within 5 Å of each other. The stronger the cross peak, the closer the two protons (or amino acids) are. Using NOESY data, in combination with the TOCSY data, it is possible to determine which amino acids are sequentially proximal to each other (*47*). In this way each set of peaks in the TOCSY spectrum can be assigned an atom name and a residue number. Once all the TOCSY peaks have been assigned an atom name and a residue number, the sequential assignment process is complete. Note that because only one type of nucleus (^1H) is being analyzed, this method is called homonuclear NMR.

When working with isotopically labeled proteins (usually $^{13}C/^{15}N$ or double-labeled), the sequential assignment process does not depend on NOESY data at all. Rather, it is possible to use a series of J-coupled 3D NMR experiments to identify individual amino acid types and to identify which pairs of amino acids are sequentially proximal to each other. These 3D NMR experiments are named for the types of nuclei they measure. Hence an HNCA experiment (*48, 49*) measures the ^1H amide shift of a residue on one axis ("H"), the ^{15}N amide shift on another axis ("N") and the $^{13}C\alpha$ shift on the third axis ("CA"). The archaebacterial proteins discussed in this work were available in both ^{15}N singly-labelled form as well as $^{15}N/^{13}C$ doubly labelled form. Therefore, both 3D ^{15}N-heteronuclear spectra and triple resonance spectra were used to complete the backbone and side-chain assignments for these proteins.

Triple resonance experiments (correlation of the frequencies of 3 different nuclei, ^1H, ^{13}C, ^{15}N) are frequently conducted on larger proteins (>10 kDa) as the problems of spectral overlap can be markedly reduced. In triple resonance experiments the magnetization is efficiently transferred through ^1J or ^2J couplings (i.e. directly via the covalent chemical bonds), therefore, the transfer times are shorter and the signal losses due to relaxation are smaller than in homonuclear experiments. A prototype triple resonance experiment HNCA (*48, 49*) is depicted in Figure 1.7. Starting at an amide proton (H) the magnetization is transferred to the directly attached nitrogen atom (N) which is measured as the first spectral dimension. Then the magnetization is transferred to the Cα nucleus which is measured in the second dimension. Afterwards, the magnetization is transferred back the same way to the amide proton, which is measured in the third (direct) dimension. In each step magnetization is transferred via strong ^1J couplings between the different nuclei. The coupling which connects the nitrogen atom with the Cα carbon of the preceeding amino acid (^2J= 7 Hz) is only marginally smaller than the coupling to the directly attached Cα atom (^1J = 11 Hz). Thus, the nitrogen atom of a given amino acid is correlated with both its own Cα and the Cα of the preceeding amino acid. Therefore, in principle, it is possible to assign the protein backbone exclusively with an HNCA experiment. However, in reality, more triple resonance experiments are needed to identify the cross signal of the preceding amino acid and to resolve degenerate resonance frequencies. An example of this is the HNCACB experiment (*50*) in which it is possible to look for very well-defined patterns to determine the amide ^{1}H shift, the amide ^{15}N shift, the ^{13}Cα shift and the ^{13}Cβ shift for each amino acid residue in the protein. The HNCACB spectrum also yields the ^{13}Cα and ^{13}Cβ shifts for the preceding residue (Figure 1.8).

A very useful triple resonance experiment for performing side-chain assignments is known as the HCCH-TOCSY (*51*). In this experiment magnetization is transferred from a sidechain (or backbone) proton to the directly attached carbon atom, by ^1J coupling to the neighboring carbon atoms and finally to their attached protons (Figure 1.9). Data from the HCCH-TOCSY experiment makes it possible to obtain chemical shifts for the ^{13}Cα as well as the side chain ^{13}C and ^{1}H atoms of each residue in the protein. By looking at the characteristic shifts of the ^{13}Cα, ^{13}Cβ and other side chain ^{13}C's obtained by the HNCACB and HCCH-TOCSY spectra one can unambiguously identify individual amino acid types. Similarly by performing the HNCO experiment

(*48*), the amide ^{1}H shift and the amide ^{15}N shift for one residue can be identified along with the ^{13}CO shift of the preceding residue (Figure 1.10). By identifying the linkages between amide shifts and the preceding ^{13}CO, ^{13}Cα and ^{13}Cβ shifts (from HNCO and HNCACB data) one can sequentially assign most of the ^{1}H, ^{13}C and ^{15}N resonances in a protein (*52, 53*). This process is called heteronuclear NMR as several different types of nuclei (^{1}H, ^{13}C, ^{15}N) are used in the assignment process.

Heteronuclear 3D spectra like ^{15}N-NOESY-HSQC (identifying through-space correlation) (*54*) and ^{15}N-TOCSY-HSQC (identifying through-bond connectivity) (*54*) can also be analyzed for assignment purposes. In these spectra, only the NOESY or TOCSY signals from those protons which are directly attached to a nitrogen atom are visible. This seperation makes 3D spectra significantly less crowded as compared to their 2D counterparts.

1.3.5 Structure Calculation in NMR

After the sequential backbone and side-chain assignments of a protein are completed, the next step in the structure determination process is to estimate proton-proton distances from signal intensities derived from NOESY experiments. NOESY experiments measure the through-space correlation between nuclei that are close in space. This correlation, termed the Nuclear Overhauser Effect or NOE (*55*), is proportional to $1/r^6$, where r is the distance between the two nuclei. The sign and intensity of the NOE also depends on the γ value of the interacting nuclei and the correlation time describing the motion of the interproton bond vector. As such, the NOEs fade quickly with distance and are not usually observed between protons that are more than 5 Å apart. However, the NOE is very sensitive to internuclear distances and therefore may be used to estimate proton-proton distances.

Secondary structural information can be reliably obtained by analyzing several parameters like short and medium range NOEs, $^3J_{HNH\alpha}$ coupling constants, amide proton exchange rates and chemical shifts. In particular, a number of short and medium range inter-proton distances (<5 Å) are fairly unique to certain secondary structural elements. For example, alpha-helices are normally characterized by short distances between backbone amide protons (d_{NN}). They are also evident between beta protons of residue i and amide protons of residue i+1 ($d_{\beta N}$), as well as between the alpha proton of residue i and the amide protons of residues i+2, i+3, and i+4. Beta-strands are characterized by short sequential distances between the alpha proton and the adjacent amide proton ($d_{\alpha N}$). As well, the formation of beta-sheet results in readily observable NOEs between protons on adjacent strands (e.g., $d_{\alpha\alpha}$ and $d_{\alpha N}$) (*42*).

$^3J_{HNH\alpha}$ coupling constants also provide very useful information about secondary structure. For example, helical and extended conformations have very different values for the ϕ angle (-60° and -120°, respectively) which result in measurable differences in $J_{HNH\alpha}$ coupling constants (*42*). Hydrogen-bonded amide protons in regular secondary structures have measurably slower exchange rates with the solvent than amides in unstructured or flexible regions. Therefore, amide proton exchange rates can provide qualitative information about the location and stability of secondary structures. For example, continuous stretches of four or more slowly exchanging amide protons can indicate the presence of a helix while alternating stretches of slowly and rapidly exchanging protons can indicate the existence of a beta-strand.

Chemical shifts can also contain useful structural information. Hα shifts as well as Cα, Cβ and CO chemical shifts for all 20 natural amino acids have been shown to have a strong correlation with secondary structure (*56*). Based on these correlations Wishart et al (*56, 57*) described a simple method for secondary structure determination by analyzing the difference between residue specific backbone ^{1}H and ^{13}C shifts in proteins and that reported for the same residue in a "random coil" conformation. More recently, chemical shifts have been used to determine approximate backbone dihedral angles by comparing observed chemical shift values to databases of known protein chemical shifts (and measured dihedral angles) of solved protein structures. Programs such as TALOS and SHIFTOR are two examples of programs which allow many dihedral angle restraints to be determined within $\pm 15^{\circ}$ (*58, 59*). By combining the secondary constants with long-range NOE information as well as the dihedral angles, one can usually determine a 3D solution structure using a variety of computational methods, namely 1) distance geometry (*60, 61*) and 2) simulated annealing (*62*).

The distance geometry method, in short, is based on a calculation of large matrices of pairwise distance constraints for each pair of atoms from all measurable (NOE) distance constraints, bond and torsion angles as well as van der Waals radii. This set of distances is then projected from the n-dimensional distance space into the three-dimensional Cartesian space. The projection or transformation allows one to determine the coordinates of all atoms of the proteins. Distance geometry is quite robust in that only sets of approximate distances are needed to generate a solution. Error and consistency checking routines along with prior information about the basics of protein structure (standard bond angles, bond lengths) also ensure that distance geometry performs well.

The simulated annealing process, on the other hand, is a molecular dynamics technique that takes place directly in Cartesian space. In this method, a starting structure is 'synthetically' heated to a high temperature (i.e. the atoms of the starting structure are given high thermal mobility) and then slowly cooled –much the same way that metal alloys are annealed. During the cooling phase the starting structure can evolve towards an energetically favourable, final structure under the influence of a molecular force field derived from the constraints. Simulated annealing generally offers a more direct and consistent approach to finding a minimum energy structure through a very complex energy landscape or hypersurface.

Regardless of the chosen technique, what is essentially done is that a random coil starting structure is first generated from empirical data. The computer program then tries to fold the starting structure in such a way, that the experimentally determined constraints are as completely satisfied as possible. In order to achieve this, each experimentally measurable parameter is assigned a pseudo-energy potential, and standard energy minimization techniques are used to reduce the overall energy. Since the protein molecule can adopt a near infinite number of conformations, it is extremely important to identify as many experimental restraints as possible to restrict the conformational search space. Typically, 60 – 100 structures are calculated in order to fully sample the allowed conformational space.

1.4 Pharmaceutical Potential of Thioredoxins and Glutaredoxins

Much of this thesis focuses on thioredoxins and glutaredoxins; therefore it is important that some background regarding their structure, function and pharmaceutical potential be provided. Thioredoxins and glutaredoxins are two classes of redox active proteins found in essentially all organisms, from bacteriophage to archaea, eubacteria, plants and animals. They take part in reactions such as redox control of transcription factors, electron transport to ribonucleotide reductase for the reduction of ribonucleosides, formation of disulfides in protein folding or defense against oxidative stress and apoptosis *(63 – 67)*.

Thioredoxins and glutaredoxins are part of the thioredoxin superfamily and are characterized by having a common fold (the thioredoxin fold - a central four-stranded beta sheet flanked by three or four alpha helices) and an active site motif consisting of Cys-Xaa-Xaa-Cys, where Xaa corresponds to any other naturally occurring amino acid (*68, 69*). The cysteines in the active site, which undergo reversible disulfide exchange reactions with their respective substrates, are always located at the N-terminus of a conserved alpha helix (*70*). In the reduced form, the N-terminal cysteine is generally solvent exposed and acts as a nucleophile whereas the more C-terminal cysteine is buried *(71)*. The difference between thioredoxin and glutaredoxin is that thioredoxin is reduced to the dithiol form by NADPH and thioredoxin reductase (the thioredoxin system) while glutaredoxin *(64)* is reduced by the tripeptide glutathione (GSH) which in turn is reduced by NADPH and glutathione reductase (the glutaredoxin system).

Thioredoxin, apart from being a normal redox component of almost all cells, is a stress-inducible protein whose expression is enhanced by various types of stressors including viral infection, exposure to UV light, X-ray irradiation and hydrogen peroxide (*72*). Thioredoxin (Trx) is a scavenger of reactive oxygen species (ROS) and recombinant Trx has been shown to have protective activity against ROS-mediated cytotoxity (*73*). ROS include free radicals such as superoxide radical anion ($O_2^{\bullet -}$), hydroxyl radical ($HO^{\bullet}$), nitric oxide (NO), singlet oxygen and other reactive metabolites such as hydrogen peroxide (H_2O_2) (*74*). ROS are mainly generated in mitochondria during oxidative phophorylation and are formed in various disease states such as inflammation, immunological reactions and ischemia (*74, 75*). ROS have putative noxious effect and have been implicated in diseases such as cancer, arteriosclerosis, autoimmune disorders, neurological degeneration, fibrosis as well as aging and wound healing (*76, 77*). Many biological systems have evolved to neutralize the toxic effects of ROS. Among these system is the thioredoxin/glutaredoxin system (*76, 77*).

Thioredoxin has been found to attenuate ischemic brain damage and prevents the thioacetamide- induced acute hepatitis in transgenic mice by scavenging radicals (*78, 79*). There is increasing experimental evidence that supports a role of oxidative stress in the pathogenesis of Alzheimer's disease (*80, 81*). Recent studies have shown that there is a decreased level of thioredoxin in Alzheimer's disease brain and treatment of cell cultures with thioredoxin lead to statistically significant concentration-dependent enhancement of cell survival (*82*). It is believed that the decreased level of thioredoxin may contribute to the increased oxidative stress and subsequent neurodegeneration observed in the brain of Alzheimer's disease (*82*). On the other hand it is also known that the level of thioredoxin in cells is elevated appreciably in many cancers (*74, 83*) and that this increased level leads to decreased sensitivity of cancers to some anti-cancer drugs (*84*). Recent studies have also shown that thioredoxin decreases the allergenicity of foods such as soy, milk and wheat in which proteins containing disulfide bonds are considered to be the major allergens (*85*, *86*).

Other non-pharmaceutical roles for thioredoxins and glutaredoxins have also been recently identified. For instance, thioredoxin is known to improve dough strength and bread quality. The evidence is in accord with the findings that thioredoxin selectively reduces intramolecular disulfide bonds of flour proteins, ie gluteins and gliadins, which then undergo a series of sulfhydryl/disulfide exchange reactions that result in the formation of new intermolecular disulfide bonds (*87*). In this manner, thioredoxin promotes the formation of a protein network, thereby enhancing the quality of the final baked product.

Other functions of thioredoxins/glutaredoxins in different organisms are listed in Table 1.2. All these actions and regulatory activities indicate that thioredoxins can play an economically important role, either as a protein pharmaceutical or as a small molecule drug target in combating many ailments that afflict man.

As mentioned earlier, our laboratory has been involved in the structural proteomic project on *M. thermoatutrophicum* (strain ΔH), working on three proteins - MTH0895, MTH0807 and MTH0776. Prior to this work, the solution structure of T4 glutaredoxin (the only protein known to have both glutaredoxin and thioredoxin activity) was determined in my supervisor's lab (*88*). When the structure of the hypothetical protein MTH0895 was determined it was found to have a glutaredoxin fold but with structure details akin to a thioredoxin. It also appears to be the smallest thioredoxin (77 amino acids) ever identified (*89*). Owing to the increasing pharmaceutical potential that thioredoxins and thioredoxin-like proteins were gaining, we decided to work on MTH0807. MTH0807 has been reported by previous workers to have some characteristics similar to both thioredoxins and glutaredoxins but it appeared to be neither a true thioredoxin nor a true glutaredoxin (*90*). We determined the structure of MTH0807 with the objective of using the structural information in conjunction with other biochemical and biophysical methods to better characterise MTH0807. This work was also undertaken to elucidate why some members of the thioredoxin superfamily behave both as a thioredoxin and a glutaredoxin while others do not. We hope that this effort will eventually aid the engineering of redox proteins with a minimal number of residues but of a particular desired activity (thioredoxin, glutaredoxin, oxidative or reductive ability). In addition to this work on thioredoxins and glutaredoxins we also determined the structure of MTH0776 (a non-thioredoxin protein with no known function) with the hope that its structure will lead us to a possible function that might eventually be useful economically to mankind.

1.5 Thesis Outline and Description of Work Done

This thesis consists of five chapters – Chapter 1 provides an introduction to the thesis and provides the necessary background for subsequent chapters. Chapter 2 focuses on describing the results of a number of biochemical tests that were used to distinguish thioredoxins from glutaredoxins. It also describes a comparison between T4 glutaredoxin and a number of thioredoxins and glutaredoxins from evolutionarily different organisms. The aim was to understand what makes T4 glutaredoxin act both as a thioredoxin and a glutaredoxin while the other proteins are not. Portions of this work have been published

(*88, 89*). Chapter 3 describes the NMR assignment and structural analysis of MTH0807 which was undertaken in an effort to better characterize its activity based on structure information. I was actively involved in this work from the expression of the protein, its purification, running of some NMR experiments (specifically, the 2D and 3D ^{15}N HSQC-NOESY and TOCSY experiments) to the complete structure determination. This work has also been published (*91*). Chapter 4 describes my work on MTH0776. I performed nearly all the work for this project, from the expression of the protein, its purification, running of all NMR experiments (with the exception of the HNCACB and T_2 experiments), processing the data, assigning the chemical shift of the protein and calculating the structure. The bioinformatics portion of the work was done by Dr. Paul Stothard while the transformation of the *E. coli* cells to produce the protein was done by Zhan Chang. The enzymatic/catalytic activity assay was performed by Alexandre Iakounine of the Best and Banting Department of Medical research, University of Toronto. The chemical shift assignment has been published (*92*) while the structure paper has been accepted for publication by the same journal. Chapter 5 is a general conclusion to the thesis and provides some thoughts on the future directions that we would like to take with this project.

1.5 References

1. Benson, D.A., Karsch-Mizarachi, I., Lipman, D.J., Ostell, J. and Wheeler, D.L. (2004) *Nucleic Acids Res.*, **32**, 23 – 26.
2. Edward, A.M., Arrowsmith, C.H. and Bertrand, des P. (2000) *Modern Drug Discovery*, **5**, 35 – 44.
3. Branca, A.M., Haberman, A.B. and Lockwood, D. (2001) Structural Proteomics: High-Throughput Approaches Fuel Drug Discovery and Developent Report **7**, Cambridge Healthtech Institute Ma., USA.
4. Watson, J.D., Laskowski, R.A. and Thornton, J.M. (2005) *Curr. Opin. Struct. Biol.*, **15**, 275 – 284.
5. Eisentein, E., Gilliland, G.L., Herzberg, O., Moult, J., Orban, J., Poljak, R.J., Banerjei, L., Richardson, D. and Howard, A.J. (2000) *Curr. Opin. Biotechnol.*, **11,** 25 – 30.
6. Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) *J. Mol. Biol.*, **247**, 536 – 540.
7. Govindarajan, S., Recabarren, R. and Goldstein, R.A. (1999) *Proteins*, **35**, 408 – 414.
8. Kim, S.H. (1998) *Nat. Struct. Biol.*, **5**, 643 – 645.
9. Gerstein, M. and Hegyi, H. (1998) *FEMS Microbial. Rev.*, **22**, 277 – 304.
10. Sali, A. (1998) *Nat. Struct. Biol.*, **5**, 1029 – 1032.
11. Sanchez, R. and Sali, A. (1998) *Proc. Natl. Acad. Sci.* USA, **95**, 13597-13602.
12. Zarembinski, T.I., Hung, L.W., Mueller-Diecmann, H-J., Kim, K-K., Yokota, H., Kim, R. and Kim, S-H. (1998) *Proc. Natl. Acad. Sci.* USA, **95**, 15189 – 15193.
13. Montelione, G.T. and Anderson, S. (1999) *Nat. Struct. Biol.*, **6**, 11 – 12.
14. Gerstein, M. (2000) *Nat. Struct. Biol.*, **7**, 960 – 963.

15. Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J.R., Booth, V., Mackereth, C.D., Saridakis, V., Ekiel, I.., Kozlov, G., Maxwell, K.L., Wu, N., McIntosh, P., Gehring, K., Kennedy, M.A., Davidson, A.R., Pai, F.E., Gerstein, M., Edwards, A.M., and Arrowsmith, C.H. (2000) *Nat. Struct. Biol.*, **7**, 903 – 909.

16. Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Le, B., Ramelot, T., Lee, G.M., Bhattacharyya, S., Gutierrez, P., Denisov, A., Lee, C-H., Cort, J.R., Kozlov, G., Liao, J., Finak, G., Chen, L., Wishart, D., Lee, W., McIntosh, L.P., Gehring, K., Kennedy, M.A., Edwards, A.M. and Arrowsmith, C.H. (2002) *Proc. Natl. Acad. Sci.* USA, **99**, 1825 – 1830.

17. Fox, G.E., Magrum, L.J., Balch, W.E., Wolfe, R.S. and Woese, C.R. (1997) *Proc. Natl. Acad. Sci.* USA, **74**, 4537 – 4541.

18. Woese, C.R. and Fox, G.E. (1997) *Proc. Natl. Acad.* Sci. USA, **74**, 5088 – 5090.

19. Doolittle, W.F. (1992) *Biochemical Society Symposium*, **58**, 1 – 6.

20. Matheson, A.T. (1992) *Biochemical Society Symposium*, **58**, 89 – 98.

21. Van Valen, L.M. and Maiorama, V.C. (1980) *Nature*, **287**, 248 – 250.

22. Madigam, M.T. and Marrs, B.L. (1997) *Scientific America*, **(Apr)**, 82 – 87.

23. Woese, C.R. (1981) *Scientific America*, **(Jun)**, 98 – 122.

24. Zeikus, J.G. and Wolfe, R.S. (1972) *J. Bacteriol.*, **109**, 707 – 713.

25. Langdahl, B.R. and Elberling, B. (1997) *Geology of Greenland survey Bulletin*, **176**, 39 – 43.

26. Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R. Vicaire, R., Wang, Y.,

Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D. and Reeve, J.N. (1997) *J. Bacteriol.*, **179**, 7135 – 7155.

27. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G. Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) *Nucleic Acids Res.*, **28**, 235 – 242.
28. Staunton, D., Owen, J. and Campbell, I.D. (2003) *Acc. Chem. Res.*, **36**, 207 – 214.
29. Wuthrich, K. (1995) *Acta Crystallogr.*, *D* **51**, 249 – 270.
30. Ishima, R. and Torchia, D.A. (2000) *Nat. Struct. Biol.*, **7**, 740 – 743.
31. Pervushin, K., Riek, R., Wider, G. and Wuthrich, K. (1997) *Proc. Nat. Acad. Sci. USA*, **94**, 12366 – 12371.
32. Bax, A. and Grzesiek, S. (1993) *Acc. Chem. Res.*, **26**, 131 – 138.
33. Zuiderweg, E.R.P. and Van Doren, S.R. (1994) *Trends Analyt. Chem.*, **13**, 24 – 36.
34. Prestegard, J.H., Tolman, J.R., Al-Hashimi, H.M. and Andrec, M. (1999) Protein structure and dynamics from field-induced residual dipolar couplings. In: Biological Magnetic Resonance: Structure Computation and Dynamics in Protein NMR (N.R. Krishna and L.J. Berliner, (Eds.), **17**, 311 – 355. Kluwer Academic-Plenum Publishers, New York.
35. Zimmerman, D.E. and Montelione, G.T. (1995) *Curr. Opin. Struct. Biol.*, **5**, 664 – 668.
36. Oschkinat, H. and Croft, D. (1994) *Meth. Enzymol.*, **239**, 308 – 318.
37. Fesik, S.W. (1993) *J. Biomol. NMR*, **3**, 261 – 269.
38. Kessler, H., Gehrke, M. and Griesinger, C. (1988) *Angewandte Chemie Internationalt (Edition in English)*, **27**, 490 – 536.
39. Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E.,

Markley, J.L. and Sykes, B.D. (1995) *J. Biomol. NMR*, **6**, 135 – 140.

40. Wishart, D.S. and Sykes, B.D. (1994) *Methods Enzymol.*, **239**, 363 – 392.

41. Karplus, M. (1959) *J. Phy. Chem.*, **30**, 11 – 15.

42. Wuthrich, K. NMR of Proteins and Nucleic Acids (1986) New York, John Wiley & Sons.

43. Dotsch, V. and Wagner, G. (1998) *Curr. Opin. Struct. Biol.*, **8**, 619 – 623.

44. Gardner, K.H. and Kay, L.E. (1998) *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 357 – 406.

45. Braunschweiler, L. and Ernst, R.R. (1983) *J. Magn. Reson.*, **53**, 521 – 526.

46. Kumar, A., Ernst, R.R. and Wuthrich K. (1980) *Biochem. Biophys. Res. Commun.*, **95**, 1 – 6.

47. Basus, V.J. (1989) *Methods Enzymol.*, **177**, 132 – 149.

48. Kay, L.E., Ikura, M., Tschudin, R. and Bax, A. (1990) *J. Magn. Reson.*, **89**, 496 – 514.

49. Grzesiek, S. and Bax, A. (1992) *J.Magn. Reson.*, **96**, 432 – 440.

50. Wittekind, M. and Mueller, L. (1993) *J. Magn. Reson.*, **101**, 206 – 210.

51. Kay, L.E., Xu, G.Y., Singer, A.U., Muhandiram, D.R. and Forman-Kay, J. *J. Magn. Reson.*, **101**, 133 – 136.

52. Ikura, M., Kay, L.E. and Bax, A. (1990) *Biochemistry*, **29**, 4659 – 4667.

53. Zheng, D., Huang, Y.J., Moseley, H.N., Xiao, R., Aramini, J., Swapna, G.V. and Montelione, G.T. (2003) *Protein Sci.*, **12**, 1232 – 1246.

54. Zhang, C. and DeLisi, C. (1998) *J. Mol. Biol.*, **284**, 1301 – 1305.

55. Noggle, J.H. and Schirmer, R.E. (1971) The Nuclear Overhauser Effects, Academic,

New York.

56. Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) *Biochemistry*, **31**, 1647 – 1651.

57. Wishart, D. S. and Sykes, B. D. (1994) *J. Biomol. NMR*, **4**, 171 – 180.

58. Cornilescu, G., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR*, **13**, 289 – 302.

59. Wishart, D.S. and Case, D.A. (2001) *Meth. Enzymol.*, **338,** 3 – 34.

60. Havel, T.F. and Wuthrich, K. (1985) *J. Mol. Biol.*, **182**, 281 – 294.

61. Williamson, M.P., Havel, T.F. and Wuthrich, K. (1985) *J. Mol. Biol.*, **182**, 295 – 315.

62. Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.*, **229**, 317 – 324.

63. Dalton, T.P., Shertzer, H.G. and Puga, A. (1999) *Annu. Rev. Pharmacol Toxicol.*, **39**, 67 – 101.

64. Holmgren, A. (1989) *J. Biol. Chem.*, **264**, 13963 – 13966.

65. Jordan, A. and Richard, P. (1998) *Annu. Rev. Biochem.*, **67**, 71 – 98.

66. Arner, E.S.J. and Holmgren, A. (2000) *Eur. J. Biochem.*, **267**, 6102 – 6109.

67. Jordan, A., Pontis, E., Aslund, F., Hellman, U., Gigert, I. and Reichard, P. (1996) *J. Biol. Chem.*, **271**, 8779 – 8785.

68. Martin, J.L. (1995) *Structure*, **3**, 245 – 250.

69. Holmgren, A. (1995) *Structure*, **3**, 239 – 243.

70. Mossner, E., Huber-Wunderlich, M. and Glockshuber, R. (1998) *Protein Sci.*, **7**, 1233 – 1244.

71. Kallis, G.B., and Holmgren, A. (1980) *J. Biol. Chem.*, **255**, 10261 – 10265.

72. Nakamurah, H., Nakamurah K. and Yodoi J. (1997) *Annu. Rev. Immunol.*, **15**, 351 – 369.

73. Nakamurah, H., Matsudla, M., Firuke, K., Kitaoka, Y., Iwata, S., Toda, K., Inamoto,

T., Yamaoka, Y., Ozawa, K. and Yodoi, J. (1994) *Immunol. Lett.*, **42**, 75 – 80.

74. Turunen, N., Karihtala, P., Mantyniemi, A., Sormunen, R., Holmgren A., Kinnula, V. and Soini, Y. (2004) *APMIS*, **112**, 123 – 132.
75. Janssen, Y.M., Van Houton, B., Boem, P.J. and Mossman, B.T. (1993) *Lab Invest.*, **69**, 261 – 274.
76. Nordberg, J. and Arnér, E.S.J (2001) *Free Radic. Biol. & Med.*, **31**, 1287 – 1312.
77. Mercurio, F. and Manning, A.M. (1999) *Oncogene*, **18**, 6163 – 6171.
78. Takagi, Y., Mitsui, A., Nishiyama, A., Nozaki, K, Sono, H., Gon, Y., Hashimoto, N. and Yodoi, J. (1999) *Proc. Natl. Acad. Sci.*, **96**, 4131 – 4136.
79. Okuyama, H., Shimahara, Y., Nakamura, H., Araya, S., Kawada, N., Yamaoka, Y. and Yodoi, J. (2004) *Comparative Hepatology*, **3**, S6
80. Markesbery, W.R. and Carney, J.M. (1999) *Brain Pathol.*, **9**, 133 – 146.
81. Markesbery, W.R. (1997) *Free Radic Biol. & Med.*, **23**, 134 – 147.
82. Lovell, M.A., Xie, C., Prasad, S.G. and Markesberry, W.R. (2000) *Free Radic. Biol. and Med.*, **28**, 418 – 427.
83. Powis, G., Mustacich, D. and Coon, A. (2000) *Free Radic Boil. & Med.*, **29**, 312 – 322.
84. Wang, J., Kobayashi, M., Sakurada, K., Imamura, M., Moriuchi, T. and Hosokawa, M. (1997) *Blood*, **89**, 2480 – 2487.
85. del Val, G., Yee, B.C., Lozano, R.M., Buchanan, B.B., Ermel, R.E., Lee, Y.M. and Frick, O.L. (1999) *J. Aller. Clin. Immunol.*, **104**, 690 – 697.
86. Buchanan, B.B. (2001) *Plant Physiol.*, **26**, 5 – 7.
87. Buchanan, B.B., Schürmann, P., Decottignies, P. and Lozano, R.M. (1994) *Arch.*

Biochem. Biophy., **314**, 257 – 260.

88. Wang, Y., Amegbey, G. and Wishart, D. (2004) *J. Biomol. NMR*, **29**, 85 – 90.

89. Bhattacharyya, S., Habibi-Nazhad, B., Amegbey, G., Slupsky, C.M., Yee, A., Arrowsmith, C. and Wishart, D. (2002) *Biochemistry*, **41**, 4760 – 4770.

90. McFarlan, S. C., Terrell, C.A., and Hogenkamp, P.C. (1992) *J. Biol. Chem.*, **267**, 10561 – 10569.

91. Amegbey, G. Y., Hassan, M., Habibi-Nazhad, B., Bhattacharyya, S. and Wishart, D.S. (2003) *Biochemistry*, **42**, 8001 – 8010.

92. Amegbey, G., Chang, Z., Stothard, P., Yee, A., Arrowsmith, C. and Wishart, D.S. (2004) *J. Biomol. NMR*, **30**, 459 – 460.

93. Holmgren, A. (1984) *Meth. Enzymol.*, **107,** 295 – 300.

94. Stewart, E.J., Åslund, F. and Beckwith, J. (1998) *EMBO J.*, **17**, 5543 – 5550.

95. Kang, S.W., Chae, H.Z., Seo, M.S., Kim, K., Baines, I.C. and Rhee, S.G. (1998) *J. Biol. Chem.*, **273**, 6297 – 6302.

96. Zhang, P., Liu, B., Kang, S.W., Seo, M.S., Rhee, S.G. and Obeid, L.M. (1997) *J. Biol. Chem.*, **272**, 30615 – 30618.

97. Chae, H.Z., Kang, S.W. and Rhee, S.G. (1999) *Meth. Enzymol.*, **300**, 219 – 226.

98. Hassouni, M.E., Chambost, J.P., Expert, D., Van Gijsegem, F. and Barras, F. (1999) *Proc. Natl Acad. Sci. USA*, **96**, 887 – 892.

99. Brot, N. and Weissbach, H. (1991) *Biofactors*, **3**, 91 – 96.

100. Russel, M. and Model, P. (1986) *J. Biol. Chem.*, **261**, 14997 – 15005.

101. Feng, J.N., Model, P. and Russel, M. (1999) *Mol. Microbiol.*, **34**, 745 – 755.

102. Lillig, C.H., Prior, A., Schwenn, J.D., Åslund, F., Ritz, D., Vlamis-Gardikas, A. and

Holmgren, A. (1999) *J. Biol. Chem.*, **274**, 7695 – 7698.

103. Schwenn, J.D., Krone, F.A. and Husmann, K. (1988) *Arch. Microbiol.*, **150**, 313 – 319.

104. Trotter, W.E. and Grant, M.C. (2002) *Molecular Biology*, **46**, 869 – 878.

105. Buchanan, B.B. (1991) *Arch. Biochem. Biophys.*, **288**, 1 – 9.

106. Wong, J.H., Balmer, Y., Cai, N., Tanaka, C.K., Vensel, W.H., Hurkman, W.J. and Buchanan, B.B. (2003) *FEBS lett.*, **547**, 151 – 156.

107. Schenk, H., Klein, M., Erdbrugger, W., Droge, W. and Schulze-Osthoff, K. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 1672 – 1676.

108. Hirota, K., Murata, M., Sachi, Y., Nakamura, H., Takeuchi, J.K.M. and Yodoi, J. (1999) *J. Biol. Chem.*, **274**, 27891 – 27897.

109. Saitoh, M., Nishitoh, H., Fujii, M., Takeda, K., Tobiume, K., Sawada, Y., Kawabata, M., Miyazono, K. and Ichijo, H. (1998) *EMBO J.*, **17**, 2596 – 2606.

110. Bertini, R., Howard, O.M., Dong, H.F., Oppenheim, J.J., Bizzarri, C., Sergi, R., Caselli, G., Pagliei, S., Romines, B., Wilshire, J.A., Mengozzi, M., Nakamura, H., Yodoi, J., Pekkari, K., Gurunath, R., Holmgren, A., Herzenberg, L.A. and Ghezzi, P. (1999) *J. Exp. Med.*, **189**, 1783 – 1789.

111. Silberstein, D.S., McDonough, S., Minkoff, M.S. and Sablinska, M.K. (1993) *J. Biol. Chem.*, **268**, 9138 – 9142.

112. Matsui, M., Oshima, M., Oshima, H., Takaku, K., Maruyama, T., Yodoi, J. and Taketo, M.M. (1996) *Dev. Biol.*, **178**, 179 – 185.

113. Goto, Y., Noda, Y., Narimoto, K., Umaoka, Y. and Mori, T. (1992) *Free Radic. Biol. & Med.*, **13**, 47 – 53.

114. Di Trapani, G., Perkins, A. and Clarke, F. (1998) *Mol. Hum. Reprod.*, **4**, 369 – 375.

115. Das, K.C., Guo, X.L. and White, C.W. (1999) *Am. J. Physiol.*, **276**, L530 - L539.

116. Hori, K., Katayama, M., Sato, N., Ishii, K., Waga, S. and Yodoi, J. (1994) *Brain Res.*, **652**, 304 – 310.

Table 1.1 Eukaryotic and Prokaryotic traits found in Archaeabacteria

Eukaryotic Traits	Bacteria traits
DNA replication machinery	No organelles
Histones	
Nucleosome-like structures	Single circular chromosome
Transcription machinery DNA-dependent RNA polymerase: Genes encoding large A', A", B' and B" and small D subunits of RNA polymerase has sequences similar to Eukaryotes than Prokaryotes. There are only Eukaryotic type E', E", H, I, K, L and N subunits in the Archaea.	Operons No introns (3 tRNA genes of M. thermoautrophicum has introns) Bacteria-type membrane transport channels
TF11B	
TATA-binding protein (TBP)	Many metabolic processes
Translation machinery	Energy production and Polysaccharide synthesis: Sequences of proteins involved in such metabolic activities are closer to the prokaryotes than the eukaryotes
Initiation factors	
Ribosomal proteins: r-protein-encoding genes encodes proteins with sequences more similar to eucaryl homologs than bacterial homologs.	Nitrogen-fixation
Elongation factors	
Poisoned by diphtheria toxin	

Table 1.2 Functions of thioredoxin in different organisms

Organism	Function and references
All organisms	**DNA synthesis and Protein disulfide reduction**. Thioredoxin is a hydrogen donor for ribonucleotide reductase (*64*) and plays a role in keeping intracellular protein disulfides generally reduced (*93, 94*)
Many Organisms	**Reduction of H_2O_2.** Many peroxidoxins, catalyzing reduction of H_2O_2 and thereby preventing oxidative stress and apoptosis induction, require reduction by thioredoxin (*95, 96, 97*) **Protein repair by methionine sulfoxide reduction**. Thioredoxin is hydrogen donor for methionine sulfoxide reductases (*98, 99*)
E. Coli phages (T7, f1, M13)	**Participates in filamentous phage assembly**. Thioredoxin is the only host *E. coli* protein required for phage assembly and export (*100, 101*)
Bacteria and Yeast	**Hydrogen donor for 3'-phosphoadenylsulfate (PAPS) reductase**. Assimilation of sulphur by sulphate to sulfite reduction (*102, 103*) **Produced in response to reductive stress in yeast**. (*104*)
Plants	**Regulation of chloroplast photosynthetic enzymes**. Photosynthesis regulation by light via ferredoxin (*105*) **Linked to many processes in endosperm of seeds** (*106*)
Mammals	**Redox regulation of transcription factors eg NFκB, AP-1**. Different transcription factors are either activated or inhibited by Trx (*107*) which may also exert different activities in the nucleus compared to the cytosol (*108*). **Regulation of Apoptosis**. Reduced thioredoxin makes a complex with ASK1 preventing downstream signalling for apoptosis (*109*) **Immunomodulation**. Extracellular thioredoxin is both a co-cytokine (*72*) and chemokine (*110*) and a truncated form stimulates eosinophils (*111*) **Pregnancy**. Intracellular and extracellular synthesis of thioredoxin from cytotrophoblasts assists implantation and establishment of pregnancy (*112, 113, 114*). **Birth.** Protection from hyperoxia at birth by induction of thioredoxin (*115*). **CNS**. Thioredoxin secreted from glial cells promotes neuronal survival at ischemia/reperfusion (*116*).

Figure 1.01 Precession of the spin angular momentum of a nucleus with non-zero spin

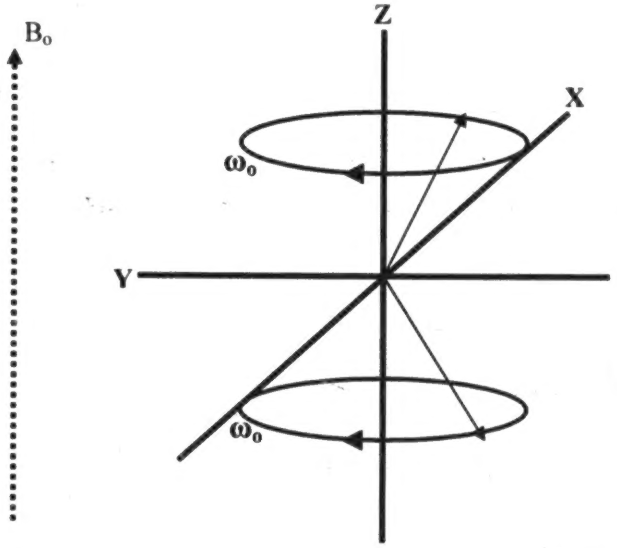


Figure 1.02 Energy level diagram for a proton with spin I = ½ in an applied magnetic field

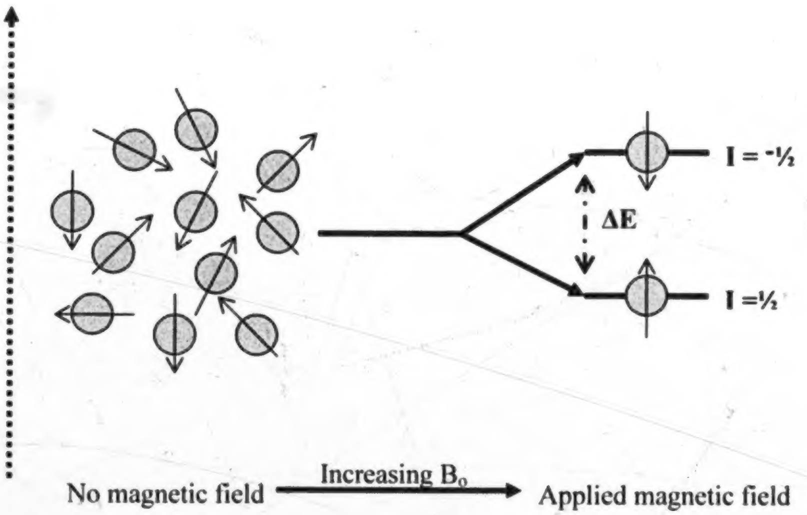

I = -½
ΔE
I = ½
No magnetic field
Increasing B₀
Applied magnetic field

Figure 1.03 Schematic illustration of the net magnetization (M) generated by population difference in an external magnetic field at equilibrium.

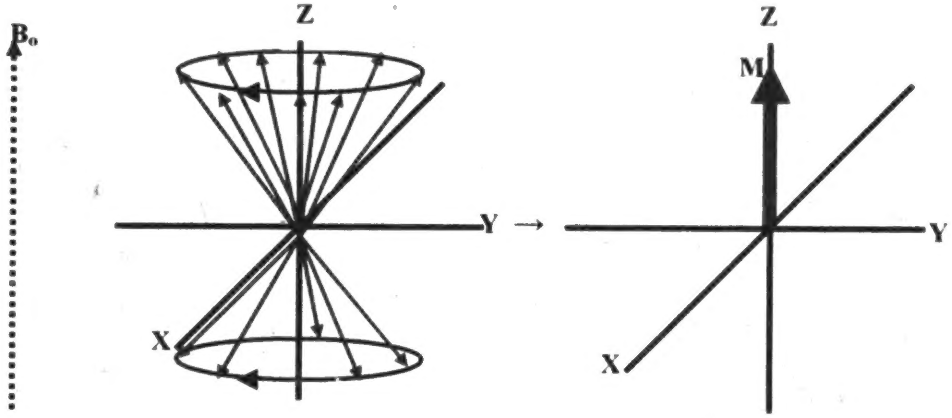


Figure 1.04 Application of a magnetic pulse 90° along the x axis with respect to the direction of B_o

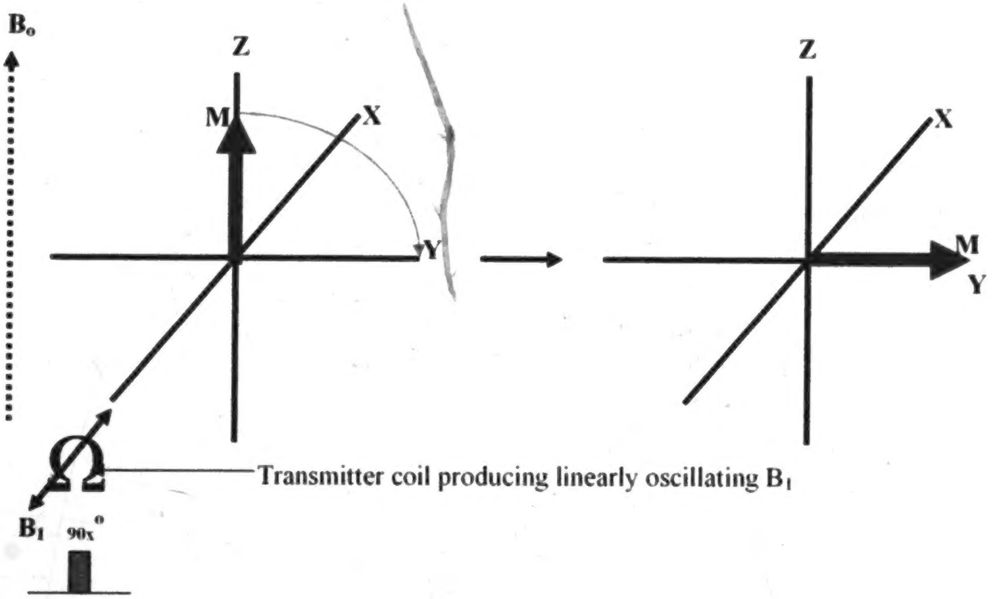

B_o
Z
M
X
Y
Z
X
M
Y
Transmitter coil producing linearly oscillating B_1
B_1
$90_x°$

Figure 1.05 After a 90° pulse along the X-axis, the net magnetization (M_o) of the system is rotated on to the Y axis completely. The magnetization (M_y) then fades (due to relaxation mechanisms) gradually with time, while the equilibrium magnetization along the Z-axis grows until the system reaches complete equilibrium again (b – e).

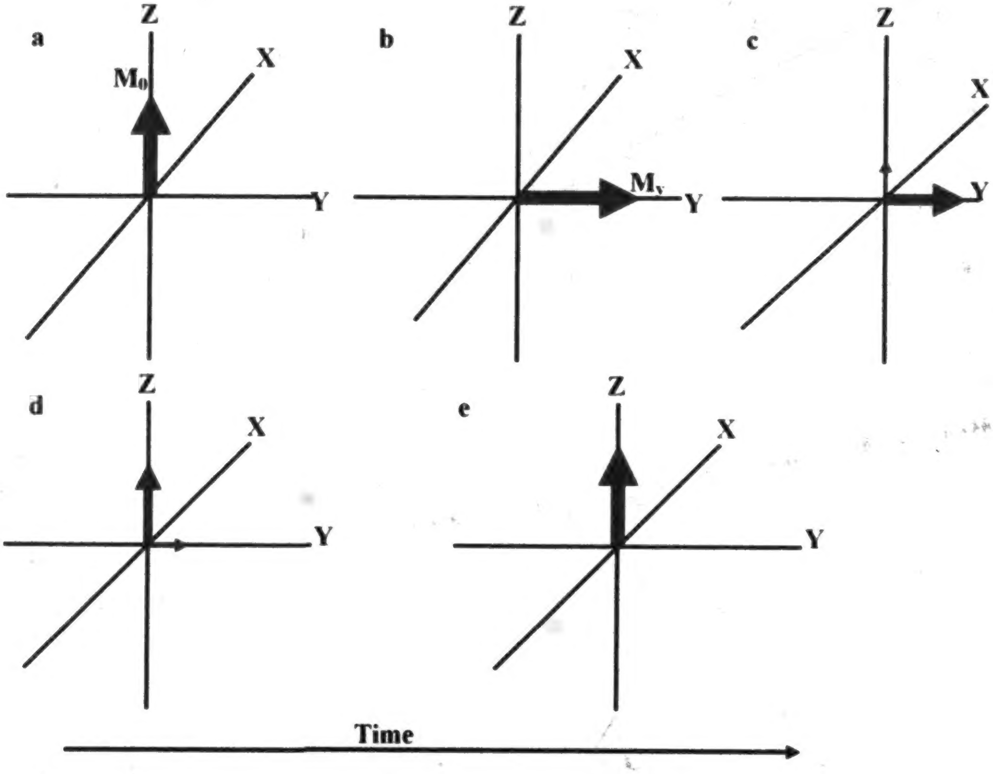


Figure 1.06 Conversion of the FID to the frequency domain by Fourier Transformation

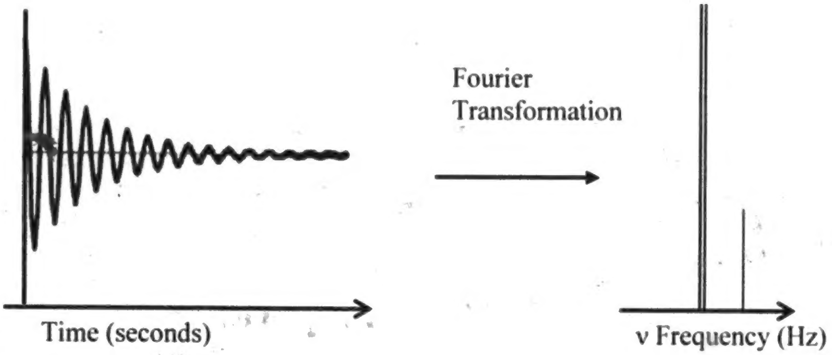

Fourier Transformation
Time (seconds)
ν Frequency (Hz)

Figure 1.07 Schematic diagram of the nuclei detected in an HNCA triple resonance experiment. Nuclei whose chemical shifts are detected are coloured grey.

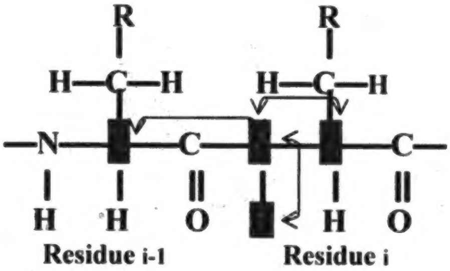

R
R
H
C
H
H
C
H
N
C
C
H
H
O
H
O
Residue i-1
Residue i

Figure 1.08 Schematic diagram of the nuclei detected in an HNCACB Triple Resonance Experiment. Nuclei whose chemical shifts are detected are coloured grey.

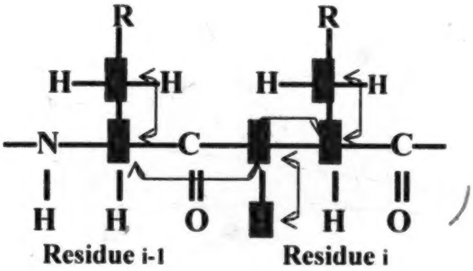
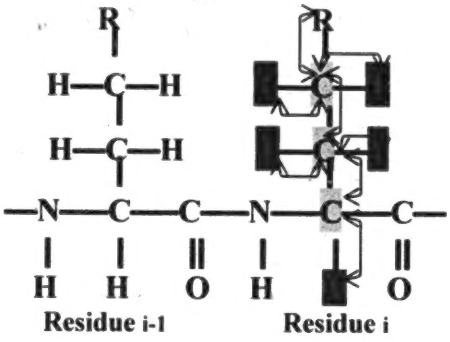
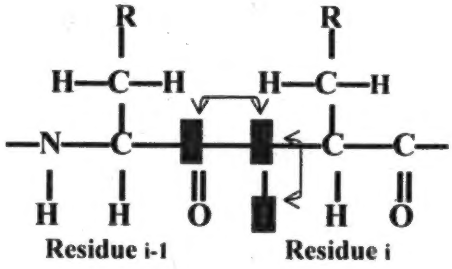

R
R
H
H
H
H
N
C
C
H
H
O
H
O
Residue i-1
Residue i

Figure 1.09 Schematic diagram of the nuclei detected in an HCCH - TOCSY Triple Resonance Experiment. Nuclei whose chemical shifts are detected are coloured grey.



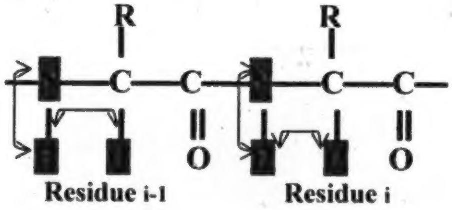
R
H—C—H
H—C—H
—N—C—C—N—C—C—
H H O H O
Residue i-1
Residue i
R
C
C
C

Figure 1.10 Schematic diagram of the nuclei detected in an HNCO triple resonance experiment. Nuclei whose chemical shifts are detected are coloured grey.



R
R
H—C—H
H—C—H
—N—C
C—C—
H H O
H O
Residue i-1
Residue i

Figure 1.11 Schematic diagram of the nuclei detected in an HNHA 3D experiment. Nuclei whose chemical shifts are detected are coloured grey.



R
R
C—C
C—C—
O
O
Residue i-1
Residue i

Figure 1.12 Schematic diagram of the nuclei detected in a C(CO)NH tripple resonance experiment. Nuclei whose chemical shifts are detected are coloured grey.

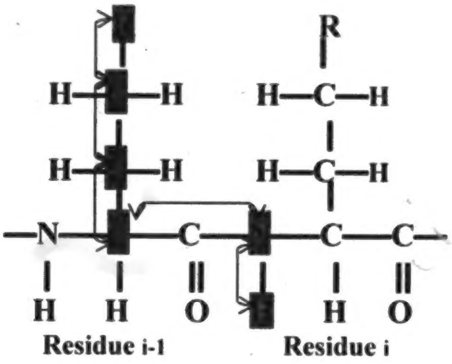
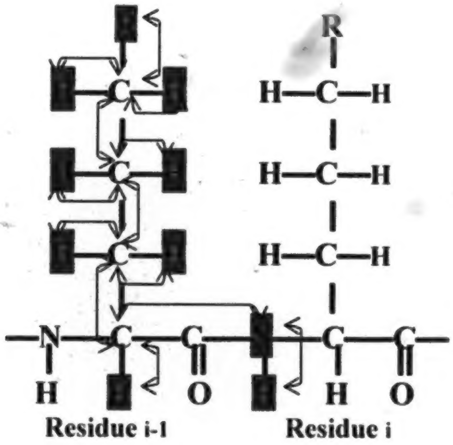

R
H—C—H
H—C—H
—N
C
C—C—
H
H
O
H
O
Residue i-1
Residue i

Figure 1.13 Schematic diagram of the nuclei detected in an H(CCO)NH tripple resonance experiment. Nuclei whose chemical shifts are detected are coloured grey.


R
H—C—H
H—C—H
H—C—H
C
C
C
—N
C
C
C—
H
O
H
O
Residue i-1
Residue i

Chapter 2

Assays and Insights into T4 Glutaredoxin's Dual Activity in Comparison to other Thioltransferases

2.1 Introduction

Glutaredoxins are small, ubiquitous, structurally conserved family of proteins typically containing fewer than one hundred amino acids. They have been identified and sequenced in essentially all classes of organisms, including bacteriophage, bacteria, yeast, plants, invertebrates and mammals (*1*). All glutaredoxins characterized to date contain a common dithiol/disulfide redox active-site motif composed of four critical residues: C-[PV]-[FYW]-C. The glutaredoxins were initially discovered as hydrogen donors for ribonucleotide reductase, a key enzyme in DNA synthesis. Other biological functions of glutaredoxins include the catalysis of dehydroascorbate to ascorbate, the reactivation of DNA-binding activity of certain nuclear factors, the acceleration of in vitro protein folding rates and the regulation and/or maintenance of HIV-1 protease activity (*1*). Glutaredoxins share many structural and functional similarities to another important member of the thioltransferase superfamily-- the thioredoxins (Trx). However, thioredoxins and glutaredoxins carry out their respective redox activities using fundamentally different mechanisms. In particular, the thioredoxin system uses NADPH, thioredoxin reductase, and thioredoxin; while the glutaredoxin system uses NADPH, glutathione reductase, glutathione, and glutaredoxin.

While sequence comparisons suggest that bacteriophage T4 glutaredoxin (T4 Grx) belongs to the glutaredoxin family, its level of sequence similarity to any other glutaredoxin rarely exceeds 20%. Furthermore, T4 glutaredoxin is the only glutaredoxin with a valine substitution of proline in the highly conserved second position of the C-P-[FYW]-C active site motif. These global and local sequence differences likely contribute to its unusual dual redox functionality as well. Recent experiments have shown that T4 Grx can be reduced not only by thioredoxin reductase but also by glutathione (*2*). Indeed, it was because of this functional hybridism that T4 glutaredoxin was originally known as bacteriophage T4 thioredoxin. It is only recently that this protein has become formally known as T4 glutaredoxin (*3*).

While several X-ray structures of oxidized T4 Grx (both native and mutant forms) have been reported (*2-6*) some years back, the solution structure of the reduced form of the wild type protein has only recently been reported (*7*). This limited structural information has limited our understanding of the structure/activity of this protein and the possible origins of its dual thioredoxin/glutaredoxin activity. More recently, the characterization of several "primitive" thioredoxins which exhibited somewhat similar thioredoxin/glutaredoxin duality (*8, 9*) to T4 Grx has led to a renewed interest in the origins and structure/function relationships of glutaredoxins and thioredoxins. In light of these new developments and in an effort to better understand the structure, activity and origin of T4 glutaredoxin we conducted a series of comparative biochemical and biophysical experiments. Specifically sequence, structure (solution structure) and biochemical activity comparisons between eukaryotic and prokaryotic thioltransferases were conducted and analyzed in detail. The results of these studies are described here, which we believe provide further insight into the origins of the dual functionality in T4 Grx.

2.2 Materials and Methods

Measurement of Thioredoxin/Glutaredoxin Activity by Ultraviolet Absorbance.

All but three of the protein reagents were purchased from Sigma (St. Louis, MI) and used without further purification. The other protein reagents (T4 glutaredoxin, MTH0807 thioredoxin and MTH0895 thioredoxin) were isolated, purified and quantified using previously described methods (*9 – 11*). To assay for thioredoxin activity, two cuvettes (blank and test) containing a 700 μL buffer solution consisting of 100 mM Tris HCl (pH 8.0), 1 mM EDTA, and 0.35 mM NADPH were prepared. 50 μL aliquots of a 0.05 mM solution of T4 glutaredoxin, *E. coli* thioredoxin (Sigma), MTH0807 thioredoxin or MTH0895 thioredoxin was added to each cuvette. After allowing the solution to equilibrate for 3 minutes, the reaction was started by the addition of 20 μL of a 0.01 mM solution of *E. coli* thioredoxin reductase (Sigma) to the test solution and 20 μL of buffer to the blank. The change in absorbance of the test solution was monitored at 340 nm (measure of NADPH oxidation to NADP based on redox potential of protein) using a Pharmacia Biotech Ultraspec 3000 UV/Vis spectrophotometer. After 25 minutes, 40 μL of a 10 mM solution of DTNB (5, 5'-dithiobis (2-nitrobenzoic acid)) was added to the test solution (40 μL of buffer to the blank) and the change in absorbance at 412 nm (measure of reduction of DTNB to TNB) was further monitored.

To assay for glutaredoxin activity, two cuvettes (blank and test) containing a solution consisting of 750 μL of 100 mM Tris HCl (pH 8.0), 1 mM EDTA, 0.35 mM NADPH, 1 mM glutathione and 0.75 nM yeast glutathione reductase (Sigma) were prepared. After 5 minutes of equilibration, a quantity (50 μL of 0.1 mM) of either MTH0807 thioredoxin, MTH0895 thioredoxin, *E. coli* thioredoxin (Sigma), *E. coli* glutaredoxin (Sigma) or T4 glutaredoxin was added to the test solution (50 μL of buffer to blank) and the change in absorbance was monitored for 25 minutes at 340 nm (measure of oxidation of glutathione) using a dual beam UV/Vis spectrophotometer (Pharmacia Ultraspec 3000).

Structural and Sequence Comparisons between T4 Grx and other Thioltransferases

We analyzed and compared the atomic structures of nine evolutionary distinct members of the thioredoxin/glutaredoxin superfamily – (*E. coli* Trx (1XOB), *E. coli* Grx 1(1EGR), *E. coli* Grx 3 (1FOV), human Trx (3TRX), human Grx (1JHB), pig liver Grx (1KTE), *M. thermoautotrophicum* Trx (Mt0807, (1NHO)), *Anabaena sp* Trx-2 (1THX) and bacteriophage T4 Grx (1DE2)).

Structures were visualised and RMSD's calculated with the Molmol program (*12*) while the residues making up the hydrophobic groove were determined from the graphical visualization tool – Swiss-PDB viewer (*13*). The primary sequence alignment based on structural alignments was obtained from the Combinatorial Extention (CE) database (*14*) and solvent accessible surface areas of the residues of the hydrophobic groove (SASA) were estimated from the program VADAR (*15*).

2.3 Results

Comparative Gluataredoxin/Thioredoxin Activity.

Figure 2a, 2b and 2c show the results of the Thioredoxin/Glutaredoxin activity assays. The thioredoxin assay is a two-step method that measures, first, the ability of the molecule to interact with thioredoxin reductase and oxidize NADPH, and second the ability of the molecule to reduce a disulfide bond. As can be seen in Figure 2b, T4 Grx is perhaps the most active thioredoxin of the group as it interacts strongly with thioredoxin reductase (albeit less so than MTH0807) and is easily oxidized by DTNB (Figure 2c). T4 glutaredoxin is followed closely by *E. coli* thioredoxin in terms of thioredoxin activity. MTH0895 Trx displays intermediate thioredoxin activity while MTH0807 shows almost no reduction ability on DTNB. In Figure 2a it is evident that T4 Grx and *E. coli* Grx display the strongest interaction with glutathioine/glutathione reductase. MTH0807 Trx exhibits about a 3-fold weaker glutaredoxin activity relative to T4 Grx while *E. coli* Trx and MTH0895 Trx shown essentially no glutaredoxin activity – as expected. The statsistical analysis (t-test) of the results is presented in appendix 1. The sequence and structural basis for the observed activity differences is discussed more fully below.

2.4 Discussion

Glutaredoxin/Thioredoxin Activity.

One of the aims of structure proteomics is to identify the function of a protein from its structure based on similarity to proteins of known function and fold. However fold similarity alone is not an indication that the protein will perform a particular function. Structure data, therefore only provides a rational lead as to the probable molecular function or specific reactivity of a hypothetical protein. Ultimately its true function has to be verified through biochemical tests.

During our work on MTH0895 (a hypothetical protein from *Methanobacterium thermoautotrophicum*), the structure was found to be similar to that of glutaredoxins while molecular docking experiments and sequence comparisons showed that it seemed more likely to behave as a thioredoxin than a glutaredoxin. In order to resolve this ambiguity, we performed the basic biochemical test that distinguishes a thioredoxin from a glutaredoxin as described by Holmgren (*16*). The principles behind this test are described in more detail in chapter 3 of this thesis. To better characterise the activity of MTH0895, we compared its activity to T4 Grx (one of only a few known proteins with both Trx and Grx activity), *Escherichia coli* Trx (which has only Trx activity), E. coli Grx (which has only Grx activity) and MTH0807, a recently characterized Trx from *Methanobacterium thermoautotrophicum*. While the general structure of MTH0895 initially indicated that it was a glutaredoxin, other indications (structural details, sequence information and molecular docking) provided strong evidence that MTH0895 was most likely a thioredoxin. The results from the thioredoxin activity assays clearly show that MTH0895 is indeed a thioredoxin. As seen in Figure 2b and 2c, MTH0895 appears to be about one-half as active as *E. coli* thioredoxin with no glutaredoxin activity exhibited. This reduced thioredoxin activity might be expected, given its short length (77 vs 108 residues), abbreviated structure (missing N-terminal β-α segment), and non-standard active site sequence (Cys-Ala-Asn-Cys vs Cys-Gly-Pro-Cys). Alternately, the lower activity may have arisen simply because the enzymatic reducing agent (thioredoxin reductase) was obtained from *E. coli* and not from *M. thermoautotrophicum*. Nevertheless, the fact that such a small glutaredoxin-like protein clearly exhibits thioredoxin activity suggests that MTH0895 and its homologues may represent a group of ancient proteins that were ancestral to both thioredoxins and glutaredoxins.

Comparison of T4 Grx with Other Glutaredoxins and Thioredoxins.

Nearly two dozen three-dimensional structures of thioredoxins and glutaredoxins have been reported over the past 25 years. To simplify the comparison process, we chose to analyze the atomic structures of nine evolutionary distinct members of the thioredoxin/glutaredoxin superfamily (*E. coli* Trx, *E. coli* Grx 1, *E. coli* Grx 3, human Trx, human Grx, pig liver Grx, *M. thermoautotrophicum* Trx (MTH0807), *Anabaena sp* Trx-2 and bacteriophage T4 Grx). Despite significant sequence differences, every study published to date has shown that these proteins share the same overall fold, with either an N-terminal $\beta_1/\alpha_1/\beta_2/\alpha_2/\beta_3$ or $\alpha_1\beta_2/\alpha_2/\beta_3$ structural unit joined to a C-terminal $\beta_4/\beta_5/\alpha_4$ unit by a loop involving the third helix (α_3). For all nine structures, the β_2, β_3, β_4, and β_5 strands form a central, continuous, twisted pleated sheet with helices α_1 and/or α_3 located on one side of the sheet and helices α_2 and α_4 located on the other side. This central β-sheet is highly conserved. Indeed, a backbone superposition of this four-stranded β-sheet for all nine highly divergent proteins (minimum pairwise sequence identity < 15%) yields a RMSD value of only 1.32 Å (Figure 2.1). This level of conservation is quite remarkable and it suggests that the central β-sheet scaffold must play a key role in the folding and stability of all members of this ubiquitous family of proteins.

In addition to the high level of structural conservation seen in the central β-sheet, there is also an equally high level of structural conservation in their active site loops. A backbone superimposition of the Cys-Xaa-Xaa-Cys redox active site for all nine solved glutaredoxin/thioredoxin molecules yields an RMSD value of just 0.36 Å (Figure 2.2). This high level of structural similarity in both thioredoxin and glutaredoxin active sites indicates just how precise the backbone geometry must be in order to confer the redox activity seen in this class of proteins. It is interesting that this structural similarity persists even though the T4 Grx active site sequence (Cys-Val-Tyr-Cys) is quite different than the usual thioredoxin (Cys-Gly-Pro-Cys) or glutaredoxin (Cys-Pro-Tyr (Phe)-Cys) active site sequences. While the active site loop is structurally well conserved, it is important to emphasize that the character and disposition of side chains around this active site is not. The nature and consequences of these differences will be discussed in more detail below.

A more quantitative comparison between T4 glutaredoxin and the other eight glutaredoxins/thioredoxins was performed using the SCOP database and its secondary structure matching system (*17*). The results are shown in Table 2.1. All structures were assessed by their Z scores, alignment length and RMSD values. As expected, the glutaredoxins scored highest with high Z values and correspondingly high levels of sequence identity. Interestingly, Grx 3 from *E. coli* shows up as the top match with a Z score of 6.4 and a sequence identity of 35%. Given that the sequence similarity between *E. coli* Grx 3 and T4 Grx is the highest among all proteins in the available sequence databanks, it suggests that the two are probably related. It also suggests that the T4 virus may have actually incorporated the *E. coli* Grx 3 gene into its genome (or vice versa). It is perhaps worthwhile noting that another redox host protein (*E. coli* Trx) has been found to be critical to the survival of a related bacterial phage (bacteriophage T7). In particular, *E. coli* Trx is known to form a complex with the phage T7 DNA polymerase to activate the enzyme (*18, 19*). It seems reasonable to suggest that bacteriophage T4, in order to overcome its dependency on bacterial host proteins for DNA synthesis, may have incorporated and modified a host redox protein (*E. coli* Grx 3) into its genome.

Understanding the Dual Activities of T4 Glutaredoxin.

A number of elegant biochemical studies have been done on T4 glutaredoxin in order to explain what makes it function both as a thioredoxin and a glutaredoxin. Notable among these are the mutational studies on amino acid residues around its active site (*18 - 20*). These studies have identified a number of residues (apart from the active site cysteines) necessary for thioredoxin activity. In particular, the following mutations: His12Ser, Lys13Ser, Lys21Gln and Pro66Ala have been found to lead to a dramatic decrease in thioredoxin activity while an Asp80Ser mutation leads to an increase in activity (*21*). A number of these residues have also been found to be important for glutaredoxin activity including His12, Pro66 and Asp80 (*22*). Many of these key residues are located in regions of the protein associated with glutathione binding (essential for glutaredoxin activity) or thioredoxin reductase docking (essential for thioredoxin activity).

As a general rule, thioredoxins typically have a large hydrophobic groove located near their active site while glutaredoxins only have a small hydrophobic pocket where glutathione binds (*23*). The hydrophobic groove in all thioredoxins is on the surface, close to the redox active cysteine such that it allows two key phenylalanine residues (Phe141 and Phe 142) in thioredoxin reductase to easily dock into this groove (*23, 24*). The groove is located in an area on the thioredoxin surface bounded by the upper halves of strands β_1 and β_2, the lower portion of strand β_3, the loop between β_2 and helix α_2 and the loop between helix α_2 and strand β_3. In an effort to better understand the importance of this groove among thioredoxins and glutaredoxins we used graphical visualization tools (Swiss-PDB viewer) to locate this hydrophobic groove in the nine selected representative members (including T4 Grx) of the thioredoxin/glutaredoxin superfamily. These other members include four glutaredoxins (*E. coli* Grx [1EGR], *E. coli* Grx 3 [1FOV] human Grx [1JHB] and pig liver Grx [1KTE]) and four thioredoxins (human Trx [3TRX], *E. coli* Trx [1XOB], *Anabaena sp.* Trx 2 [1THX] and *M. thermoautotrophicum* (MTH0807) Trx [1NHO]). Structure-based sequence alignments of these proteins are shown in Figure 2.3

After identifying the thioredoxin reductase binding surface (or analogous surface on glutaredoxins), we then used VADAR (*15*) to quantitatively estimate the solvent accessible surface area (SASA) of each of the hydrophobic residues making up this hydrophobic groove. Our results, listed in Table 2.2 show that the SASA of this hydrophobic surface in T4 Grx is large (442 $Å^2$) and comparable to those of other thioredoxins including *E. coli* Trx (465 $Å^2$), MTH0807 Trx (408 $Å^2$), Human Trx (402 $Å^2$) and *Anabaena* Trx 2 (400 $Å^2$). In contrast, the SASA of the hydrophobic residues in the surface for most glutaredoxins was much less with *E. coli* Grx 1, *E. coli* Grx 3, Pig liver Grx and Human Grx having accessible areas of just 168 $Å^2$, 60 $Å^2$, 72 $Å^2$ and 65 $Å^2$. Note that the total SASA of the two critical phenylalanine residues in thioredoxin reductase is in excess of 210 $Å^2$. Clearly T4 Grx has a large enough hydrophobic groove to accept thioredoxin reductase, while other glutaredoxins do not.

In addition to a larger-than-expected hydrophobic groove in T4 Grx, this protein also has an unusually large (for a glutaredoxin) loop between helix α_2 and strand β_3. This loop has been reported to be a common determinant in the recognition and binding of all thioredoxins to thioredoxin reductase (*24*) and has been shown to dock into a complementary groove on the surface of the thioredoxin reductase molecule. As can be seen in Figure 2.3, the length of this loop in T4 Grx is longer than in other known glutaredoxins and comparable to those of thioredoxins. This loop was also found to be highly flexible in a previous NMR study which is typical of many binding loops used in protein-protein interactions (*25*). The size and flexibility of this loop, in combination with the unusually large size of T4 Grx's hydrophobic groove near the active site surface might account for the observed binding of T4 glutaredoxin to thioredoxin reductase and for its observed thioredoxin activity.

T4 Grx is not alone in its ability to function as both a glutaredoxin and a thioredoxin. By looking at the structural similarities of other "dual-function" redox proteins we tried to gain some further insight into the key residues or features that make these proteins function as they do. As seen in Figure 2a, our glutaredoxin activity assays clearly show that T4 Grx and *E. coli* Grx have the highest level of glutaredoxin activity while thioredoxins such as MTH0895 Trx and *E. coli* Trx have essentially no glutaredoxin activity. However, MTH0807 Trx, a small archaebacterial thioredoxin with a classic glutaredoxin active site sequence, does show some measurable glutaredoxin activity – roughly one third of that measured for T4 Grx. Another thioredoxin, *Anabaena* Trx 2, has also been reported to be reduced by glutathione with an activity level approximately one fifth that of T4 Grx (*8*). This capacity for all three thioredoxins to function as glutaredoxins appears to be related to the presence of a viable glutathione binding site. All known glutaredoxins have a glutathione binding pocket on their surface which is largely absent in "classic" thioredoxins. For instance, the binding pocket in T4 Grx consists of Tyr7, His12, Tyr16, Thr64, Met65, Pro66 and Asp80 (*20*). A possible glutathione binding pocket on MTH0807 Trx, analogous to that of T4 Grx, is Thr9, Tyr15, Val56, Pro57 and Ser72. The glutathione binding pocket (peptide binding pocket) for *Anabaena* Trx 2 has also been characterized and appears to consist of Pro64, Lys68,Gly74, Val75, Pro76, and Ser94 (*26*).

The downward trend in glutaredoxin activity going from T4 Grx, to MTH0807 Trx and then to *Anabaena* Trx 2 appears to agree well with results reported from mutational studies done on T4 Grx (*20*). Specifically, two positively charged residues, His12 and Lys21 (T4 Grx numbering) were found to be important for activity in T4 Grx. However, both are absent in MTH0807 Trx and *Anabaena* Trx 2. Additionally, while MTH0807 has a classic glutaredoxin active site sequence (Cys- Pro - Tyr-Cys), *Anabaena* Trx 2 lacks tyrosine in its active site, having instead Cys-Gly-Pro-Cys active site sequence. Previous studies have shown that tyrosine in the active site is important in glutathione binding (*20*). Thus the order of glutaredoxin activity is expected to be T4 Grx > MTH0807 Trx > *Anabaena* Trx 2, as was observed.

Additional comparisons between MTH0807 Trx, *Anabaena* Trx 2 and T4 Grx in terms of their thioredoxin activity are also instructive. All known thioredoxins have a Lys or an Arg residue located on the C-terminal side of the active site cysteine. It has been suggested that this positively charged residue near the active site probably helps in stabilizing the N-terminal thiolate anion of the reduced form (*26*). Interestingly T4 Grx (a glutaredoxin, with thioredoxin activity) also has a lysine in the same position. How important is this positively charged residue for thioredoxin activity? Sequence comparisons shown in Figure 2.3 indicate that MTH0807 Trx and *Anabaena* Trx 2 are the only two thioredoxins without an equivalent positive residue near the active site in the conserved helix α_1. Interestingly, both MTH0807 Trx and *Anabaena* Trx 2 have been reported to exhibit thioredoxin-like activities, including an ability to reduce insulin and an ability to provide reducing equivalents to ribonucleotide reductase *(8, 9, 27)*. However, neither *Anabaena* Trx 2 nor MTH0807 Trx is able to reduce DTNB in a significant way – a common test for thioredoxin activity. Therefore we would suggest that a basic amino acid residue on the C-terminal side of the active site is probably essential for the binding and reduction of small molecules such as DTNB but not for large protein substrates such as ribonucleotide reductase.

Despite the apparent functional similarities, [illegible] appears that T4 Grx is quite distinct from these other dual-function redox proteins (MTH0807 Trx and *Anabaena* Trx) in that T4 Grx exhibits remarkably good glutaredoxin and thioredoxin activity while the other two proteins only exhibit limited activity (as either thioredoxins or glutaredoxins). It may be that MTH0807 Trx developed its glutaredoxin activity "accidentally" and has been under no selective pressure to enhance or diminish this activity. Certainly archaebacteria have no need for glutathione or for glutaredoxin activity. On the other hand, it appears that T4 Grx must have been under selective pressure to produce a protein with strong glutaredoxin and thioredoxin activity as this protein exhibits or retains nearly all of the critical sequence and structural features required for both activities.

2.5 Conclusion

In this study we described the results of both comparative biochemical assays and comparative sequence/structure assessments involving phage T4 glutaredoxin with a range of eubacterial, archaebacterial and eukaryotic thioredoxins and glutaredoxins. These include four glutaredoxins (*E. coli* Grx, *E. coli* Grx 3, human Grx and pig liver Grx) along with five thioredoxins (human Trx, *E. coli* Trx, *Anabaena sp.* Trx 2, *M. thermoautotrophicum* (MTH0895) Trx and *M. thermoautotrophicum* (MTH0807). We have shown through our biochemical tests that even though MTH0895 and MTH0807 show a classic glutaredoxin fold, they actually behave functionally as a thioredoxin. This further augments the fact that the fold of a protein only gives us a lead as to its possible biochemical function. Structure details in addition to biochemical testing are the only sure proof of function.

Comparative analyses of the solution structure of T4 Grx to the other glutaredoxins and thioredoxins indicated that the large hydrophobic groove on its surface (close to the redox active site) as well as the highly flexible nature of the loop joining helix α_2 and strand β_3 might be the main reason T4 Grx is able to interact with thioredoxin reductase (thereby allowing it to function as a thioredoxin). These results could shed further light onto the reasons and origins of the dual redox character of T4 glutaredoxin and might suggest ways of engineering multiple activities into other thioredoxins or glutaredoxins. It would be interesting to see what the results would be if residues corresponding to His12 and Lys21 (T4 Grx numbering) in MTH0807 are mutated to His and Lys respectively.

2.6 References

1. Holmgren, A., Arner, E.S.J. and Berndt, K.D. (1999) Glutaredoxin: In the Encyclopedia of Molecular Biology, Creigthon, T.E. (Ed), John Wiley and Sons, Inc. New York, pp 1020 – 1023.
2. Nikkola, M., Gleason, F.K. and Eklund, H. (1993) *J. Biol. Chem.*, **268**, 3845 – 3849.
3. Nikkola, M., Gleason, F.K., Saarinen, M., Joelson, T., Bjornberg, O. and Eklund, H. (1991) *J. Biol. Chem.*, **266**, 16105 – 16112.
4. Soderberg, B.O., Sjoberg, B.M., Sonnerstam, U. and Branden, C.I. (1978) *Proc. Natl. Acad. Sci. U. S. A.*, **75**, 5827 – 5830.
5. Eklund, H., Ingelman, M., Soderberg, B.O., Uhlin, T., Nordlund, P., Nikkola, M., Sonnertam, U., Joelson, T. and Petratos, K. (1992) *J. Mol. Biol.*, **228**, 596 – 618.
6. Ingelman, M., Nordlund, P. and Eklund, H. (1995) *FEBS Lett.*, **370**, 209 – 211.
7. Wang, Y., Amegbey, G. and Wishart, D. (2004) *J. Biomol. NMR*, **29**, 85 – 90.
8. Gleason, F. K. (1992) *J. Bacteriol.*, **174**, 2592 – 2598.
9. Amegbey, G. Y., Hassan, M., Habibi-Nazhad, B., Bhattacharyya, S. and Wishart, D.S. (2003) *Biochemistry*, **42**, 8001 – 8010.
10. Bhattacharyya, S., Habibi-Nazhad, B., Amegbey, G., Slupsky, C.M., Yee, A., Arrowsmith, C. and Wishart, D. (2002) *Biochemistry*, **41**, 4760 – 4770.
11. LeMaster, D.M. and Richards, F.M. (1988) *Biochemistry*, **27**, 142 – 150.
12. Koradi, R., Billeter, M. and Wuthrich, K. (1996) *J. Mol. Graphics*, **14**, 51 – 55.
13. Guex, N. and Peitsch, M.C. (1997) *Electrophoresis*, **18**, 2714 – 2723.
14. Shindyalov, I. N. and Bourne, P. E. (1998) *Protein Eng.*, **11**, 739-747.
15. Willard, L., Anuj, R., Zhang, H., Monzavi, H., Boyko, R., Sykes, B.D. and Wishart,

D.S. (2003) *Nucl. Acids Res.*, **31**, 3316 – 3319.

16. Holmgren, A. (1979) *J. Biol. Chem.*, **254**, 3664 – 3671.
17. LoConte, L., Brenner, S.E., Hubbard, T.J.P., Chothia, C. and Murzin, A. (2002) *Nucl. Acid Res.*, **30**, 264 – 267.
18. Mark, D.F. and Richardson, C.C. (1976) *Proc. Natl. Acad. Sci.* USA, **73**, 780 – 784.
19. Tabor, S., Huber, H.E. and Richarson, C.C. (1987) *J. Biol. Chem.*, **262**, 16212 – 16223.
20. Nikkola, M., Gleason, F.K., Saarinen, M., Joelson, T., Bjornberg, O. and Eklund, H. (1991) *J. Biol. Chem.*, **266**, 16105 – 16112.
21. Nikkola, M., Gleason, F.K. and Eklund, H. (1993) *J. Biol. Chem.*, **268**, 3845 – 3849.
22. Navaro, J.A., Gleason, F.K., Cusanovich, M.A., Fuchs, J.A., Meyer, T.E. and Tollin, G. (1990) *Biochemistry*, **30**, 2192 – 2195.
23. Stehr, M., Schneider, G., Aslund, F., Holmgren, A. and Lindquist, Y. (2001) *J. Biol. Chem.*, **276**, 35836 – 35841.
24. Lennon, B.W., Williams Jr., C. H. and Ludwig, M.L. (2000) *Science*, **289**, 1190 – 1194.
25. Wilson, I.A. and Stanfield, R.L. (1993) *Curr. Opin. Struct. Biol.*, **3**, 113 – 118.
26. Saarinen, M., Gleason, F.K. and Eklund, H. (1995) *Structure*, **3**, 1097 – 1108.
27. McFarlan, S.C., Terrell, C.A. and Hogenkamp, P.C. (1992) *J. Biol. Chem.*, **267**, 10561 – 10569.

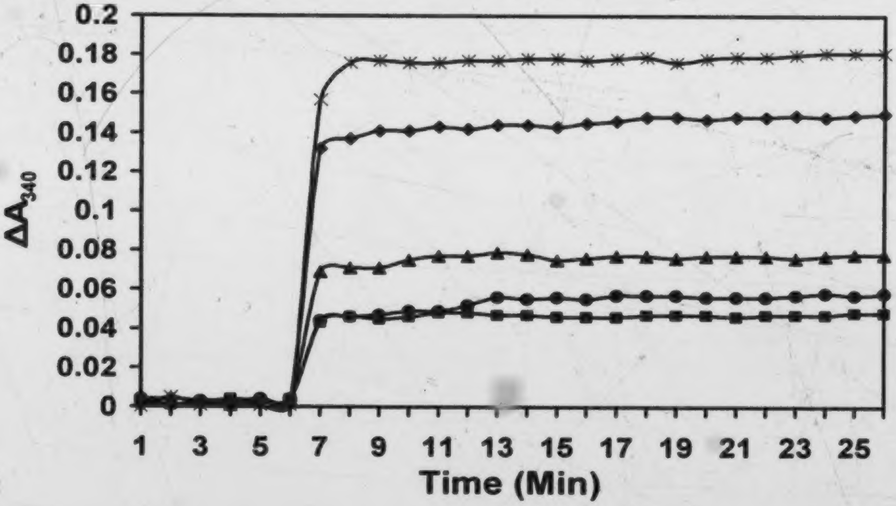
Table2.1 Structure comparison of T4 Grx (1DE2) and other thioredoxins and glutaredoxins derived from the PDB database using the SCOP structure alignment database:

Protein	**Z Score**	**RMSD**	**Aligned length**	**Seq. length**	**% Identity of aligned length**	**No of gaps**	**PDB code**
E. coli Grx 3	6.40	1.91	71	82	35	4	1FOV
Sus scrofa Grx	5.20	2.45	70	105	29	4	1KTE
E. coli Grx 1	4.10	2.51	74	100	19	6	1EGR
Homo sapiens Grx	3.70	2.65	69	105	28	6	1JHB
M. thermoautrophicum Trx	3.50	3.26	64	85	19	7	1NHO
Anabaena sp. Trx 2	3.20	2.53	65	108	8	7	1THX
E. coli Trx	2.90	2.00	60	108	8	8	1XOB
Homo sapiens Trx	1.70	3.32	64	105	6	7	3TRX

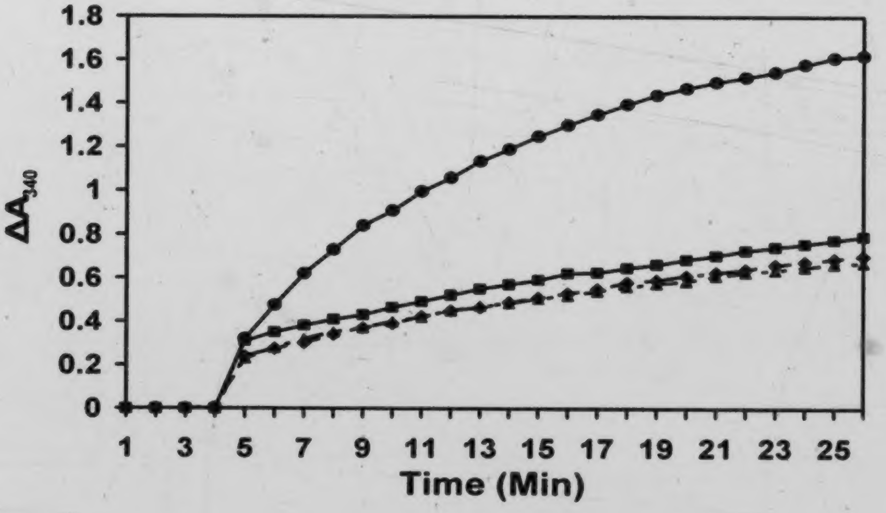
Table 2.2 Solvent accessible surface area (SASA) of the hydrophobic groove on the surface of thioredoxins and glutaredoxin into which two phenylalanine residues of thioredoxin reductase docks

Name and PDB code of Molecule	**Residues making up the Hydrophobic groove**	**SASA ($Å^2$)**
E. coli Trx (1XOB)	Leu79, Ile75, Gly74, Ile72, Ile60, Trp31, Ile23	465
T4 Grx (1DE2)	Phe69, Pro66, Met65, Leu63, Gly62, Leu55,Leu51, Ile48, Phe43, Ile36, Ile34	442
M. thermoautrophicum Tr (1NHO)	Val65, Val64, Gly63, Ile61, Pro57, Val56, Ile41, Ile39	408
Homo sapiens Trx (3TRX)	Gly87, Val86, Phe80, Pro75, Val71, Cys69, Ala66, Val59, Leu55, Trp31, Ala29	402
Anabaena sp Trx 2 (1THX	Leu87, Ile86, Pro76, Val75, Gly74, Val72, Ile60, Trp31	400
E. coli Grx 1 (1EGR)	Gly70, Ile69, Val64, Phe63, Pro60, Val59, Ile	168
E. coli Grx 3 (1FOV)	Gly63, Ile62, Phe56, pro53, Ile33	72
Sus scrofa Grx (1KTE)	Cys78, Phe73, Gly75, Pro70, Val69	65
Homo sapiens Grx (1JHB)	Cys79, Gly76, Ile75, Phe74, Val70	60

Figure 2 Measurement of thioredoxin and glutaredoxin activity via ultraviolet absorbance (a) Glutaredoxin activity kinetics. The change in absorbance (Δ_{340}, absolute values) for T4 Grx (-*-*-*-), *E. coli* Grx (-◆-◆-◆-), *E. coli* Trx (-●-●-●-), Mt0895 Trx (-■-■-■-) and Mt0807 Trx (-▲-▲-▲-). The change in absorbance is proportional to the amount of glutathione oxidised hence a measure of the glutaredoxin activity



(b) Thioredoxin activity kinetics (oxidation of NADPH). The change in absorbance (Δ_{340}, absolute values) for T4 Grx (-■-■-■-), *E. coli* Trx (-◆-◆-◆-), Mt895 Trx (-▲-▲-▲-) and Mt0807 (-●-●-●-).The change in absorbance is proportional to the amount of NADPH oxidised (depending on the redox potential of the protein) and hence a measure of the thioredoxin/thioredoxin reductase interaction



c) Thioredoxin activity kinetics (reduction of DTNB). The change in absorbance (Δ_{412}) for T4 Grx (-■-■-■-), *E. coli* Trx (-◆-◆-◆-), Mt895 Trx (-▲-▲-▲-) and Mt0807 Trx (-●-●-●-). The change in absorbance is proportional to the amount of DTNB reduced hence a measure of the reductive ability of the protein

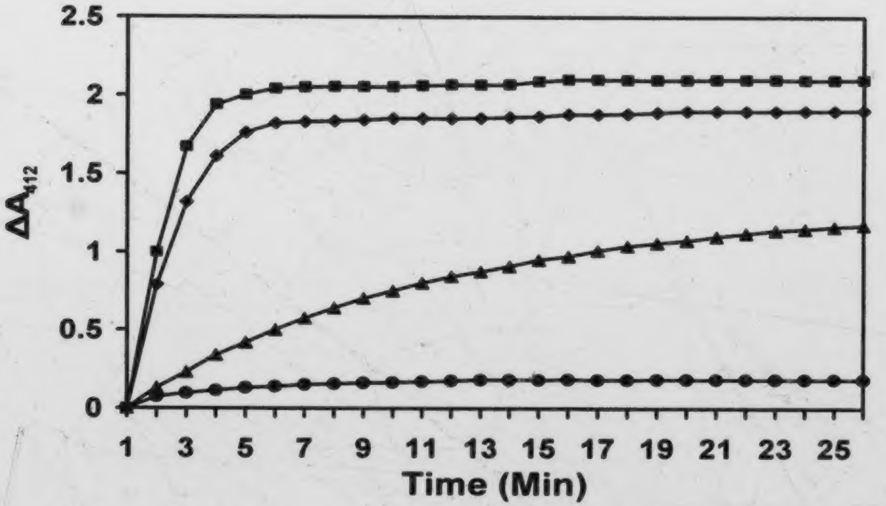


Figure 2.1 Backbone superposition of the four - stranded β-sheet for nine highly divergent proteins of the thioredoxin superfamily: T4 Grx (1DE2), E. coli Grx 1 (1EGR), Homo sapiens Grx (1JHB), Sus scrofa Grx (1KTE), E. coli Grx 3 (1FOV), Homo sapiens Trx (3TRX), E. coli Trx (1XOB), Anabaena sp. Trx-2 (1THX) and *M. thermoautotrophicum* Trx (MTH0807) (1NHO).

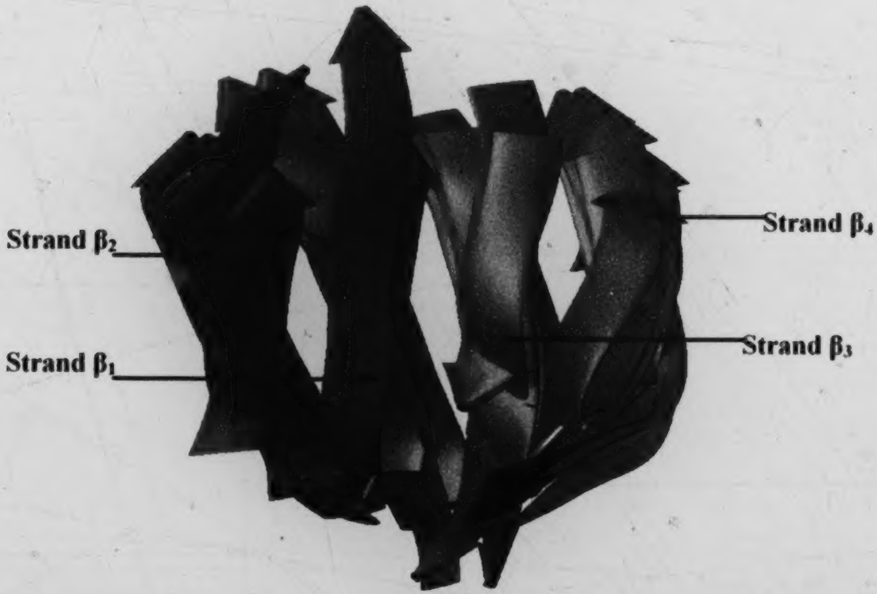


Figure 2.2 Backbone superposition of the active site Cys-Xaa-Xaa-Cys sequence for nine highly divergent proteins of the thioredoxin superfamily: T4 Grx (1DE2), E. coli Grx 1 (1EGR), Homo sapiens Grx (1JHB), Sus scrofa Grx (1KTE), E. coli Grx 3 (1FOV), Homo sapiens Trx (3TRX), E. coli Trx (1XOB), Anabaena sp. Trx-2 (1THX) and *M. thermoautotrophicum* Trx (MTH0807) (1NHO).

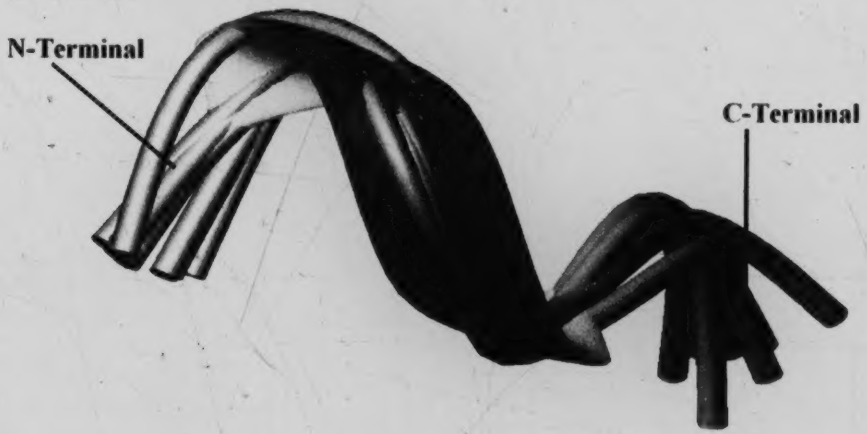
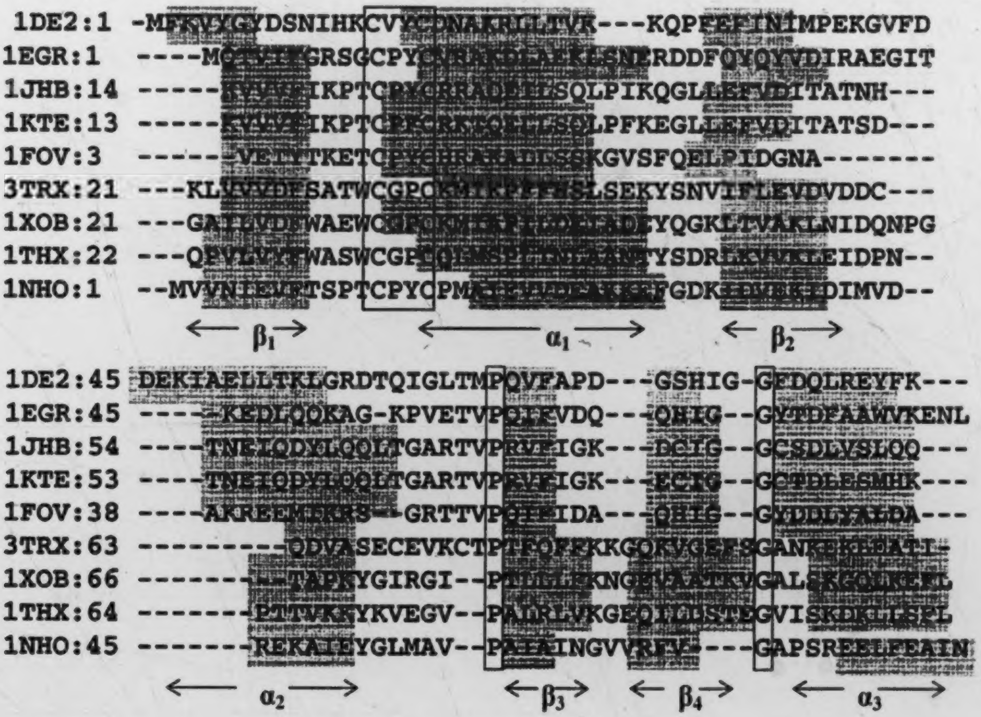


Figure 2.3 Structure-based sequence alignment obtained from the CE database of: T4 Grx (1DE2), E. coli Grx 1 (1EGR), Homo sapiens Grx (1JHB), Sus scrofa Grx (1KTE), E. coli Grx 3 (1FOV), Homo sapiens Trx (3TRX), E. coli Trx (1XOB), Anabaena sp. Trx-2 (1THX) and *M. thermoautotrophicum* Trx (MTH0807) (1NHO). The β-strands and α-helices are marked grey. The active site sequences, conserved cis proline and the conserved glycine residue are enclosed in boxes.


1DE2:1 -MFKVYGYDSNIHKCVYCDNAKRLLTVK---KQPFEFINIMPEKGVFD
1EGR:1 ----MQTVIFGRSGCPYCVRAKDLAEKLSNERDDFQYQYVDIRAEGIT
1JHB:14 -----KVVVFIKPTCPYCRRAQEILSQLPIKQGLLEFVDITATNH---
1KTE:13 -----KVVVFIKPTCPYCRKTQELLSQLPFKEGLLEFVDITATSD---
1FOV:3 ------VEIYTKETCPYCHRAKALLSSKGVSFQELPIDGNA-------
3TRX:21 ---KLVVVDFSATWCGPCKMIKPFFHSLSEKYSNVIFLEVDVDDC---
1XOB:21 ---GAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPG
1THX:22 ---QPVLVYFWASWCGPCQLMSPLINLAANTYSDRLKVVKLEIDPN--
1NHO:1 --MVVNIEVFTSPTCPYCPMAIEVVDEAKKEFGDKIDVEKIDIMVD--
β1
α1
β2
1DE2:45 DEKIAELLTKLGRDTQIGLTMPQVFAPD---GSHIG-GFDQLREYFK---
1EGR:45 -----KEDLQQKAG-KPVETVPQIFVDQ---QHIG--GYTDFAAWVKENL
1JHB:54 ----TNEIQDYLQQLTGARTVPRVFIGK---DCIG--GCSDLVSLQQ---
1KTE:53 ----TNEIQDYLQQLTGARTVPRVFIGK---ECIG--GCTDLESMHK---
1FOV:38 ----AKREEMIKRS--GRTTVPQIFIDA---QHIG--GCDDLYALDA---
3TRX:63 ---------QDVASECEVKCTPTFQFFKKGQKVGEFSGANKEKLEATI-
1XOB:66 ---------TAPKYGIRGI--PTLLLFKNGEVAATKVGALSKGQLKEFL
1THX:64 -------PTTVKKYKVEGV--PALRLVKGEQILDSTEGVISKDKLLSFL
1NHO:45 -------REKAIEYGLMAV--PAIAINGVVRFV----GAPSREELFEAIN
α2
β3
β4
α3

Chapter 3

Structural and Functional Characterisation of MTH0807

3.1 Introduction

MTH0807 is the formal protein abbreviation for the protein coded by gene 0807 of *Methanobacterium thermoautotrophicum* (strain ΔH), a hyperthermophilic archaeon. MTH0807 from *Methanobacterium thermoautotrophicum* (strain Marburg) has been shown by McFarlan and co-workers (*1*) to be a redox protein with an active site sequence similar to known glutaredoxins (-Cys-Pro-Tyr-Cys-) but with some properties inconsistent with either a glutaredoxin or a thioredoxin. Specifically, they reported that the protein does not function as a glutathione-disulfide oxido-reductase in the presence of glutathione/glutathione reductase, thereby suggesting it was not a glutaredoxin. However, they also found that even though it exhibited some thioredoxin-like properties, it was not a substrate for the thioredoxin reductase, thereby suggesting that it was not a true thioredoxin either.

Recall that thioredoxins and glutaredoxins are characterized by having a common fold (the thioredoxin fold - a central four-stranded beta sheet flanked by three or four alpha helices) and an active site motif consisting of Cys-Xaa-Xaa-Cys, where Xaa corresponds to any other naturally occurring amino acid (*2, 3*). The cysteines in the active site, which undergo reversible disulfide exchange reactions with their respective substrates, are located at the N-terminus of a conserved alpha helix (*4*). In the reduced form, the N-terminal cysteine is generally solvent exposed and acts as a nucleophile whereas the more C-terminal cysteine is buried (*5*). The difference between thioredoxin and glutaredoxin is that thioredoxin is reduced to the dithiol form by NADPH and thioredoxin reductase (the thioredoxin system) while glutaredoxin (*6*) is reduced by the tripeptide glutathione (GSH) which in turn is reduced by NADPH and glutathione reductase, (the glutaredoxin system). Thioredoxins and glutaredoxins can also be distinguished by their lack of cross reactivity with antibodies raised specifically against each protein (*1*).

Recent whole genome analyses of various hyperthermophilic organisms (*7* - *13*) as well as NrdH proteins isolated from mesophilic organisms (*14, 15*) have revealed that there are classes of small redox proteins which have sequences and folds similar to glutaredoxins but which have true thioredoxin-like activities (*16* - *18*). Most interestingly, studies by Bhattacharyya et al. (*17*) have revealed the presence of another paralogous thioredoxin (MTH0895) in *M. thermoautotrophicum* which shares 28% sequence identity to MTH0807. Additionally, work by Lee et al. (*16*), has demonstrated true thioredoxin activity for a small redox protein from *Methanococcus jannaschii* (Mj0307) which has 51% sequence identity to MTH0807.

In order to clarify the role of MTH0807 and its relationship (both structural and functional) to its paralogue MTH0895, we undertook additional biochemical/biophysical tests on MTH0807. We have also determined its solution structure via NMR to explore the structure-function relationship between MTH0807 and MTH0895. Our findings show that MTH0807, despite having a classic glutaredoxin active site sequence (-Cys-Pro-Tyr-Cys-) and a classic glutaredoxin fold, is actually a true thioredoxin. Sequence comparisons, detailed structural analyses of the charge distribution around the active site as well as enzymatic assays (which employed a more sensitive two step redox activity assay than used by previous workers (*1*)); clearly indicate that MTH0807 is similar to other known thioredoxins from archaeabacteria. Our data also show that MTH0807 is more oxidizing than its paralogue MTH0895, even though the two proteins have very similar structures and appear to use the same reducing system.

3.2 Materials and Methods

Protein Expression, Purification and Sample Preparation.

The MTH0807 gene was provided by Professor Cheryl Arrowsmith of the division of Molecular and Structural Biology, Ontario Cancer Institute (Toronto, Canada). The gene was sub-cloned into a pET15b vector (Novagen) and transformed into a BL21 (DE3) strain of *E. coli* for expression. The vector expresses the protein with an N-terminal hexa-histidine tag followed by a thrombin cleavage site (figure 3.1). Protein expression was accomplished by growing the transformed *E. coli* cells in a minimal (M9) growth medium. Uniform isotopic enrichment with ^{15}N or $^{15}N/^{13}C$ was accomplished by substituting the NH_4Cl and/or glucose in the M9 medium with $^{15}NH_4Cl$ and [U-^{13}C]-glucose (Martek Labs, Richmond VA) respectively. Cells were grown at 37 °C in shaker flasks to an optical density of 0.6 and then induced with isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. The cells were allowed to grow for an additional eight hours before harvesting and centrifugation. Cell pellets were resuspended in a lysis buffer (50 mM $NaHPO_4$, 300 mM NaCl, 10 mM imidazole, pH 8.0). Using 5ml of the buffer per gram (wet weight) of cell pellet, lysozyme was added to a concentration of 10 μg/mL. The mixture was then incubated in dry ice for 30 minutes and thawed in a water bath at 37 °C for 30 minutes. This freeze-thaw process was repeated three times. The lysed cells were then centrifuged at 15,000 rpm for 90 minutes at 4 °C. The supernatant was isolated and RNase and Dnase 1 were added to a concentration of 10 μg/mL. The mixture was incubated at room temperature for 10 minutes and centrifuged at 15,000 rpm for 45 minutes. The protein was then purified in the native condition using the nickel nitrilotriacetic acid-agarose resin (Ni-NTA) column (Qiagen) as described elsewhere (*19*). The histidine tag was not cleaved from the protein as this type of tag doe does not interfere with the spectral properties of the protein (*20, 21*). The isolation and purification process yielded 20mg of histidine tagged protein per litre of minimal (M9) growth medium as determined by UV absorbance at 280 nm using an extinction coefficient of 3230 M^{-1} cm^{-1}.

NMR Spectroscopy

NMR samples were prepared by dissolving about 6 mg of protein in 500 μL of a buffer solution (pH 6.0) made up of 50 mM NaH_2PO_4, 100 mM NaCl, 50 μL of D_2O, 1 mM DSS and 10 μL of a 3% solution of sodium azide. NMR experiments were recorded at 25 °C on a Varian Inova 500 MHz spectrometer equipped with a 5 mm triple resonance and pulse gradient accessories. Two-dimensional ^{1}H-NOESY, (mixing times of 80 ms and 150 ms), 2-D ^{1}H-TOCSY (mixing time of 60 ms) experiments were acquired on the unlabelled sample (*22, 23*). Additionally, ^{1}H-^{15}N HSQC, ^{1}H-^{15}N NOESY-HSQC (mixing time 80 ms and 150 ms), and ^{1}H-^{15}N TOCSY-HSQC (mixing times of 50 ms and 100 ms) experiments were collected on the ^{15}N labeled sample (*23, 24*). In addition, HNCO (*25*) and HNCACB (*26*) data were collected on the ^{15}N/^{13}C labeled sample. All spectra were processed with an in-house spectral processing program

(PROSIGN) which is a menu driven package written specifically for the Varian VNMR environment.

Assignments and Experimental Restraints

Sequential chemical shift assignments for MTH0807 were obtained by identifying ^{13}Cα(i) to ^{13}Cα (i-1) and ^{13}Cβ(i) to ^{13}Cβ(i-1) connectivities from the HNCACB spectrum. The backbone chemical shift assignment was then completed using data from HNCO and ^{1}H-^{15}N TOCSY-HSQC spectra. Side chain ^{1}H assignments of each amino acid was then added using the ^{1}H-^{15}N HSQC (figure 3.2) and spin systems obtained from the ^{1}H-^{15}N TOCSY-HSQC spectra. Stereo-specific assignments of ^{1}Hβ protons were based on the intensities of ^{1}HN-^{1}Hβ and ^{1}Hα-^{1}Hβcross-peaks in the ^{1}H-^{15}N NOESY-HSQC and 2-D ^{1}H-NOESY spectra. The methyl groups of Val and Leu were assigned stereo-specifically based on the intensity of ^{1}HN-^{1}Hγ, ^{1}Hα–^{1}Hγ cross peaks. NOE distance restraints were obtained from the ^{1}H-^{15}N NOESY-HSQC and 2-D ^{1}H-NOESY spectra. NOE peak intensities were obtained from the volume integration of well resolved cross peaks. Assigned NOE restraints were classified into three distance ranges: 1.8-2.7 Å, 1.8-3.5 Å and 1.8-5.0 Å corresponding to strong, medium, and weak NOE intensities respectively. Torsion angle constraints were predicted using an in house program called SHIFTOR (*27*) which calculates both the phi and psi torsion angles from observed ^{1}HN, ^{1}Hα, ^{13}Cα, ^{13}Cβ and ^{13}CO chemical shifts. Hydrogen bond restraints ($d_{O\text{-}HN}$ = 1.8-2.7 Å and $d_{O\text{-}N}$ = 2.8-3.7 Å) were identified from the pattern of sequential and inter-strand NOEs involving amide and alpha protons, D_2O exchange ^{15}N HSQC experiment, and from the chemical shift index (*28*).

Structure Calculations

Structures for MTH0807 were calculated using X-PLOR version 3.851 (*29*). Sixty random structures were generated using X-PLOR's standard simulated annealing protocol for NMR structure determination (*30, 31*). Only NOE derived internuclear distance constraints representing the secondary structure elements were used (*32*). These random structures were then regularized using the distance geometry/simulated annealing protocol (dgsa.inp) (*31, 33, 34*) of X-PLOR with additional sets of NOE constraints. The regularization stage was repeated after correcting NOE distance violation (usually by extending their upper limits) after visually reassessing the corresponding NOE spectral intensities. Typically these problem NOEs were borderline cases falling between strong/medium or medium/weak NOE intensity. Dihedral angle and the hydrogen bond restraints were then introduced and the structures were further refined using X-PLOR's simulated annealing refinement (refine.inp) protocol for NMR structure determination (*33*). A total of 997 restraints (comprising of 214 intra-residue NOEs, 659 inter-residue NOEs (345 sequential, 187 medium range and 127 long range), 82 dihedral angle restraints and 42 hydrogen bonds) were used for the structure calculations. The final set of 20 structures was selected based on the acceptance criteria that no inter-proton distance violation could be more than 0.5 Å, no torsion angle, bond angle or improper dihedral angle restraint violation could be greater than 5.0° and no bond violation could be greater than 0.05Å. The dihedral angle and NOE-(distance)-refined structures were further refined using the mini_shift_coup.inp protocol of X-PLOR (*35 - 37*). This protocol refines the structures based on NOE distance, dihedral angle and chemical shift restraints. The refined structures were analyzed with PROCHECK-NMR (*38*) and

VADAR (*39*). MOLMOL (*40*) was used to visualize all the structures (figure 3.3 and figure 3.4) and to calculate the RMSD values.

Measurement of Thiol Ionization in MTH0807 by Ultraviolet Absorbance

Thiol Ionization constants for MTH0807 were measured using the protocol of Dyson et al. (*41*). Fresh samples of MTH0807 (1.0 mL of a 1 mM sample in 0.1 M potassium phosphate, 0.1 mM EDTA, pH 7.0 under argon) was reduced using 1.0 mL of a 10 mM DTT solution. Excess DTT was removed by dialysis against the same buffer under an argon atmosphere. After dialysis, the protein sample was diluted to a concentration of about 30 μM in a 0.1 mM EDTA, and 100 mM potassium phosphate buffer (pH 6.0). Thiol ionization was monitored by measuring the protein absorbance at 240 nm on a Pharmacia Biotech Ultraspec 3000 UV/Vis spectrophotometer. Small aliquots of 1.0 M NaOH or 2.0 M HCl were added to adjust the pH up or down. After each addition the change in protein concentration was noted at 280 nm. The concentration of MTH0807 was calculated at 280 nm using an extinction coefficient of 3230 $M^{-1}cm^{-1}$.

Measurement of Thioredoxin/Glutaredoxin Activity.

Thioredoxin/Glutaredoxin activity measurements were performed using the method described by Holmgren (*42*). To assay for glutaredoxin activity, a solution containing 750 μL of 100 mM Tris HCl (pH 8.0) with 1 mM EDTA, 0.35 mM NADPH, 1 mM glutathione and 0.75 nM yeast glutathione reductase (Sigma, St Louis MI) was prepared. After 5 minutes of equilibration, a quantity (20 μL of 0.2 mM) of either

MTH0807, MTH0895, *E. coli* thioredoxin (Sigma), *E. coli* glutaredoxin (Sigma) or T4 glutaredoxin (*43*) was added to the reaction mixture and the change in absorbance monitored at 340nm. To assay for thioredoxin activity, a solution containing 750 μL of 100 mM Tris HCl (pH 8.0) with 1 mM EDTA, 0.35 mM NADPH and 0.02 mM of MTH0807, MTH895, T4 glutaredoxin or *E. coli* thioredoxin was prepared. After allowing the solution to equilibrate for 3 minutes, 20 μL of a 0.01 mM solution of *E. coli* thioredoxin reductase (Sigma) was added and the oxidation of NADPH monitored at 340 nm using a Pharmacia Biotech Ultraspec 3000 UV/Vis spectrophotometer. After 25 minutes, 40 μL of a 10 mM solution of DTNB (5, 5'-dithiobis (2-nitrobenzoic acid)) was added and the change in absorbance at 412 nm was further monitored. All reagents purchased from Sigma (St. Louis, MI) were used without further purification while the concentrations of MTH0807, MTH0895 and T4 Grx were determined by measuring the UV absorbance at 280 nm with extinction coefficients of 3230, 1615 and 6612 $M^{-1}cm^{-1}$ respectively (*17, 43*).

3.3 Results

Solution Structure

Statistical parameters for the ensemble of 20 calculated structures are presented in Table 3.1. All the structures (Figure 3.4), show good covalent geometry as indicated by a low RMSD from idealized values and by low NOE, dihedral angle and van der Waals energies. The refined structures were analyzed with PROCHECK-NMR (*38*) which indicated that for all 20 structures, 99.0% of the main chain (ϕ and ψ) angles fall into the allowed regions of the Ramachandran map. With the exception of the active site loop, (residues 12-16) and the C-terminus (residues 84-85), all portions of the MTH0807 structure are well-defined by NMR standards. The complete set of 20 MTH0807 structures has been deposited with the Protein Data Bank (PDB accession: 1NHO). The complete set of ^{1}H, ^{13}C and ^{15}N chemical shifts has been deposited with the BioMagResBank (BMRB accession number: 5622). Visual inspection of the structure shows that MTH0807 is made up of a four stranded β-sheet sandwiched between two helices on one side and a third on the other (figure 3.3). The β-strands have a $\uparrow\beta_4\downarrow\beta_3\uparrow\beta_1\uparrow\beta_2$ orientation with helices 1 and 3 packing against each other on one side of the β-sheet and helix 2 on the other. The four β-strands in MTH0807 include residues 3-8 (strand β_1), 34-39 (strand β_2), 58-61 (strand β_3) and 65-68 (strand β_4). The first helix (α-helix$_1$) is the longest and runs from residue 19-30. The third helix (α-helix$_3$), is on the same side of the sheets as α-helix$_1$, and is composed of residues 74-84. The second helix (residues 45-50) exhibits properties of a 3_{10} helix and connects strand β_2 to β_3. Depite having 'helical' chemical shifts, the 3_{10} helix was found to have characteristic $H\alpha(i) - HN(i + 2)$ NOE connectivities along its length was marked by an absence of $H\alpha(i) - HN(i + 4)$ NOE connectivities. The presence of a 3_{10} helix was also confirmed by the secondary structure analysis of the final protein structures with VADAR (*39*) and MOLMOL (*40*). Another structural feature characteristic of the thioredoxin/glutaredoxin fold include a cis-peptide bond preceeding strand β_3 at Pro 57 (*2*) as confirmed by the $H\alpha(i) - H\alpha(i + 1)$ and $HN(i) - H\alpha(i + 1)$ NOE connectivities (*32*). Additionally, the side chain of Phe 8 is found to pack into a hydrophobic pocket formed by the N-terminus of α-helix$_2$ and it is highly conserved in all disulfide oxido-reductases as either a tyrosine or a phenylalanine residue (*44*). The topological arrangement of the secondary structural elements (β-α-β-α-β-β-α) and the overall 3D structure clearly indicates that MTH0807 has a glutaredoxin-like fold (*6*) as has been found for other members of the archaeabacterial redox protein family (*16, 17*).

Sequence and Structural Comparison with Other Thioredoxins and Glutaredoxins.

Primary sequence analysis was conducted on MTH0807 using PSI-BLAST (*45*). Results obtained from the database searches clearly show that MTH0807 belongs to the thioredoxin/glutaredoxin superfamily. It shows relatively high sequence identity with orthologous members in the Archaea family (ranging from between 51%- 28%) and also with eubacterial thioredoxins/glutaredoxins. However, the level of sequence identity of MTH0807 to the well characerized members of the thioredoxin superfamily rarely exceeded 20%. Primary sequence alignment based on structural alignments obtained from the Combinatorial Extension (*46*) database was done on MTH0807 using representative structures from the thioredoxin/glutaredoxin superfamily. The structures included: thioredoxin H from *Chlamydomonas reinhardtii* (1TOF), thioredoxin from *E. coli* (1XOB), Human thioredoxin (4TRX), thioredoxin from *M. jannaschii* (1FO5), thioredoxin from *M. thermoautotrophicum* (1ILO), glutaredoxin from *E. coli* (1EGR) and human glutaredoxin (1JHB). The result is shown in Figure 3.5 and it indicates that even though the sequence identities might be low, the secondary structure however is very much conserved. These results also show that the active site sequence in MTH0807 is a classic (-Cys-Pro-Tyr-Cys-) glutaredoxin active-site sequence.

A qualitative structural comparison was carried out between MTH0807 and other representative members of the thioredoxin/glutaredoxin superfamily including: 1) T4 glutaredoxin (1DE1); 2) Thioredoxin (E. coli, 1XOB); 3) Glutaredoxin-1 (E. coli, 1EG0); 4) Human glutaredoxin (1JHB); 5) Human thioredoxin-like (1GH2); 6) Thioredoxin MTH0895 (1ILO) and 7) Thioredoxin Mj0307 (1F05). This detailed visual analysis revealed obvious similarities among all the structures as expected for the glutaredoxin/thioredoxin superfamily. One difference of note between the structure of MTH0807 and the standard glutaredoxin fold is the connection between β_4 and α_3. In the archaeabacterial redox proteins, this connection are made up of between 4-6 residues whereas this junction in *E. coli* glutaredoxin-1 (and all mesophilic prokaryotic or eukaryotic glutaredoxin structures) is rather short, typically consisting of a single glycine residue (*44*). Therefore this loop is more similar to that of *E. coli* thioredoxin. Also helix$_2$ is considerably shorter in archaebacterial redox proteins than in glutaredoxins but comparable to that of thioredoxin from *E. coli*. It was also observed that the connection between β_2 and α_2 is much shorter than in *E. coli* glutaredoxin but comparable to that of *E. coli* thioredoxin. A hydrophobic pocket at the surface and the loop between α_2 and β_3 is a common determinant in all other thiol redoxins, for the recognition and binding of thioredoxin to thioredoxin reductase. This loop occupies a complementary groove on the surface of thioredoxin reductase upon binding (*47*). In MTH0807, this loop (residues 51-56) occupies a similar position and orientation when MTH0807 and the other thioredoxin-like proteins are superimposed - much more so than when it is superimposed with *E. coli* glutaredoxin. At the C-terminal end of β_4 is glycine residue (Gly69 in MTH0807) which is highly conserved in the thioredoxin superfamily (*48*). This residue has been shown to be important in binding of thioredoxin to other molecules such as T7 DNA polymerase, thioredoxin reductase (*49*) or in the assembly of filamentous phages

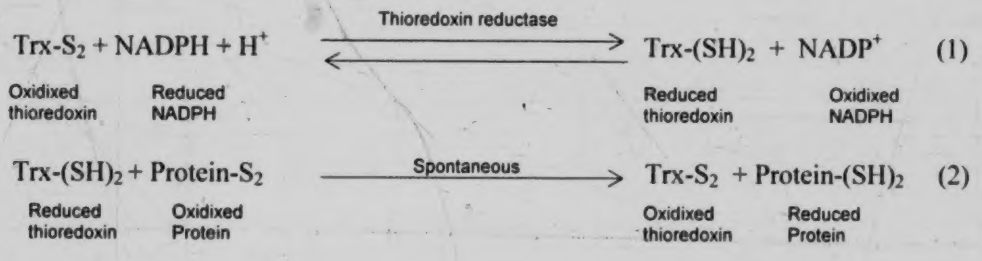
(*50*). One unique difference between the structure of MTH0807 and that of the other known oxido-reductase proteins is that helix$_2$ is a 3_{10} helix in MTH0807 whereas it is an alpha helix in the others. More quantitative comparison of the structure of MTH0807 with other known thioredoxins/glutaredoxins was also achieved by submitting the representative structure of MTH0807 to be searched against the PDB, Dali (*51, 52*), SCOP (*53*) and CE (*46*) databases. The results are listed in Table 3.2 and commented on in the discussion section.

Thiol Ionization by UV Spectroscopy

The thiol Ionization measurement of the pKa is based on the principle that the thiol (-SH) and the thiolate ion ($-S^-$) absorbs UV light to different extents at a wavelength of 240 nm with a difference in extinction coefficient of about $4000 M^{-1}cm^{-1}$ (*54 - 56*). This extinction coefficient difference is sufficiently high to prevent facile monitoring of the thiol ionization without interference from peptide absorption bands at 200-220nm or aromatic absorption at 260nm. Figure 3.6 shows the result of the thiol ionization test conducted on MTH0807. A titration is clearly visible with a midpoint at ~ pH 6.2 and a change in extinction coeffecient of about 7900 $M^{-1}cm^{-1}$ between pH 5.90 and 6.40 corresponding to two thiols titrating at the same pH. Fitting this value to the Henderson-Hasselbach equation gives an approximate pKa of 6.23 for both Cys13 and Cys16 in the active site of the enzyme.

Thioredoxin Activity

Thioredoxin reductase specifically reduces oxidized thioredoxin to reduced thioredoxin using NADPH. The reduced thioredoxin forms a powerful protein disulfide reductase and spontaneously reduces other oxidized proteins. This is expressed in equations 1 and 2 below and it describes the principle behind all thioredoxin/thioredoxin reductase assays (*57*).



$$\underset{\text{Oxidixed thioredoxin}}{\text{Trx-S}_2} + \underset{\text{Reduced NADPH}}{\text{NADPH}} + \text{H}^+ \xrightleftharpoons{\text{Thioredoxin reductase}} \underset{\text{Reduced thioredoxin}}{\text{Trx-(SH)}_2} + \underset{\text{Oxidixed NADPH}}{\text{NADP}^+} \qquad (1)$$

$$\underset{\text{Reduced thioredoxin}}{\text{Trx-(SH)}_2} + \underset{\text{Oxidixed Protein}}{\text{Protein-S}_2} \xrightarrow{\text{Spontaneous}} \underset{\text{Oxidixed thioredoxin}}{\text{Trx-S}_2} + \underset{\text{Reduced Protein}}{\text{Protein-(SH)}_2} \qquad (2)$$

Thioredoxin reductase is a FAD containing enzyme and functions by transferring reducing equivalents to a disulfide bond in the active site of a redox protein via FAD. The thioredoxin reductase subunits contains FAD and NADPH binding domains and studies suggests that thioredoxin reductase must undergo a large conformational change during enzyme catalysis in order for oxidized thioredoxin to be reduced by the dithiol of the reduced enzyme (*58*). *E. coli* thioredoxin reductase is specific for its homologous oxidized thioredoxin molecule but it also shows reactivity with other thioredoxin molecules from other organisms (*57*). The results of the thioredoxin activity assay for MTH0807 indicate that the first part (Figure 3.7b) of the two-step reaction is occurring in all the proteins used in the test. This indicates that thioredoxin reductase is reducing the proteins and MTH0807 is by far the most oxidizing protein of the protein molecules tested. The second step (Figure 3.7c) seems to be non-existent for MTH0807 while T4 glutaredoxin and *E. coli* thioredoxin showed the highest activity and MTH0895 shows only moderate activity. The statistical significance of the comparisons of these activities is presented in appendix 1. It is worth noting here that T4 glutaredoxin is the only redox protein reported to have both strong glutaredoxin and thioredoxin activity (*1, 59*). Our results suggest that even though MTH0807 is showing a very strong interaction with *E. coli* thioredoxin reductase, it lacks the spontaneous reduction activity observed for many thioredoxin-like molecules in its class. MTH0807 is, therefore, more of an oxidizing protein than a reducing one. This is manifested by its low pKa value and may also be related to the very reducing environmental niche that *M. thermoautotrophicum* occupies (*4, 16, 60*).

Glutaredoxin Acivity.

Glutaredoxin activity is shown by Holmgren (*61*) as a two step reaction as indicated by equations 3 and 4 below.

$$\underset{\text{Oxidized Glutaredoxin}}{\text{Glutaredoxin-S}_2} + \underset{\text{Reduced Glutathione}}{\text{2GSH}} \longrightarrow \underset{\text{Reduced Glutaredoxin}}{\text{Glutaredoxin-(SH)}_2} + \underset{\text{Oxidized Glutathione}}{\text{GSSG}} \tag{3}$$

$$\underset{\text{Oxidized Glutathione}}{\text{GSSG}} + \underset{\text{Reduced NADPH}}{\text{NADPH}} + \text{H}^+ \xrightarrow{\text{Glutathione Reductase}} \underset{\text{Reduced Glutathione}}{\text{2GSH}} + \underset{\text{Oxidized NADPH}}{\text{NADP}^+} \tag{4}$$

The glutaredoxin activity test therefore, should depend on the ability of the redox protein to bind and oxidize glutathione. Figure 3.7a shows the results of this test for T4 glutaredoxin, *E. coli* glutaredixin, *E. coli* thioredoxin, MTH0895 and MTH0807. The glutaredoxin activity observed for MTH0807 is more than that observed for MTH0895 (a recently characterized thioredoxin from *M. thermoautotrophicum*) and *E. coli* thioredoxin. It has been demonstrated that *E. coli* thioredoxin is reduced extremely slowly by NADPH, GSH and glutathione reductase (*59*). The activity of MTH0807 is however much lower than those observed for T4 glutaredoxin and *E. coli* glutaredoxin. The weak glutaredoxin activity observed for MTH0807 could be due to some hydrophobic binding to glutathione and the high oxidative properties that it exhibits.

3.4 Discussion

Sequence and Structure Comparison.

Sequence comparison between MTH0807 and 'standard' thioredoxins/glutaredoxins (Figure 3.5) indicate that only a few residues are fully conserved. As a rule, the active site sequence –Trp-Cys-Gly-Pro-Cys is highly conserved in the thioredoxins where as the glutaredoxins have the conserved Cys-Pro-Tyr-Cys sequence in their active site (*17*). On the basis of its sequence alone, MTH0807 would be expected to behave like a glutaredoxin. Furthermore, the structure of MTH0807 clearly shows that it has a glutaredoxin-like fold. Typically a glutaredoxin fold is similar to a thioredoxin-like fold except that the N-terminal residues (forming a β-α secondary structure) are truncated. However, more detailed structural comparisons with other members of the thioredoxin superfamily indicate that the general fold, even though similar to that of glutaredoxin, has several characteristics more common to a thioredoxin. Furthermore, biochemical tests suggest MTH0807 functions much more like a thioredoxin.

Submission of the representative structure to search for structure neighbors in the DALI, CE, SCOP and PDB databases yielded very useful results. Table 3.2 lists the summary of the top matches from a SCOP database search. These same matches were also among the top matches in the CE, PDB and Dali databases. The structures were assessed by their Z scores, alignment length and RMSD values. Interestingly, the top matches from all these searches were from Archaebacteria with MTH0807's paralogue, MTH0895 topping the list and Mj0307 (its othologue) in second place. These were followed by the glutaredoxins and then bacteriophage T4 glutaredoxin. At the bottom of the list were either the thioredoxins or glutathione transferases. These results indicate that even though MTH0807 is a true thioredoxin, its overall fold is closer to the glutaredoxins than the thioredoxins, suggesting (in addition to other evidence discussed below) that it may be ancestral to the glutaredoxins.

While these fold comparisons are useful for understanding the possible evolution or phylogeny of archaebacterial thioredoxins, it is also important to emphasize that more detailed structural comparisons can help explain MTH0807's observed activity. In particular, the similarities between MTH0807 and other known thioredoxins in terms of the length and orientation of helix$_2$, the loop connecting helix$_2$ and strand β_3, the presence of a glycine residue at the C-terminal end of strand β_4 and the hydrophobic pocket on the surface of the molecule – all of these features are key determinants for thioredoxin reductase binding (*49*), which is observed for MTH0807. Additional structural comparisons between MTH0807 and known glutaredoxins show that MTH0807 lacks the positively charged residues at the opposite sides of its redox-active disulfide bridge which are thought to be essential to glutaredoxin activity.

Thiol Ionization by UV spectroscopy

The thiol ionization test performed clearly gave results that are consistent with those observed for *E. coli* thioredoxin and with other archaeabacterial thioredoxins. The structure of MTH0807 obtained by NMR shows that the active site cysteines (Cys13 and Cys16) are in a loop between strand-β_1 and α-helix $_1$ (figure 3.4). Cysteine 13 is exposed while cysteine 16 is buried. As observed for *E. coli* thioredoxin, the –SH group of Cys13 is expected to be neutralized prior to Cys16 as the pH is increased (*5*). Thus for *E. coli* thioredoxin, the pKa's for the exposed and buried cysteines have been reported to be 7.1 and 9.9 respectively (*43*). In glutaredoxins the pKa of the exposed cysteine is typically below 5 (those of thioredoxins are above 6.5) while the buried cysteine typically has a pKa above 9 as in thioredoxins (*62 - 64*). It has also been found that the ionization of the active site thiols is significantly affected by the presence of other charged groups in the molecule. In *E. coli*, these groups are Asp26 and Lys57 and when either one or both residues are mutated in *E. coli*, the two thiols either titrate simultaneously with the same pKa value (*41, 65*) or titrate at very close/similar pKa's (*66*). Aspartic acid 26 is completely buried and in close proximity to the active site (the closest approach being the carboxyl group of Asp26 to the sulfur of Cys35 at 5.9 Å in reduced thioredoxin and 5.6 Å in oxidized thioredoxin (*67*)). The function of Lys57 is not completely clear but it is thought to be a possible hydrogen bond donor or salt bridge partner for the Asp 26 carboxyl group (*41*). The pKa of thioredoxins/glutaredoxins is also influenced by the residues in between the active site thiols (*16*).

Structural analysis of the MTH0807 protein shows that neither of these groups (Asp or Lys) is proximal to the sulfur of Cys16. Thus a two-step endpoint with very similar pKa's or a single pKa for both thiols would be expected (as was observed). The pKa value of 6.23 for both thiols is more typical for the exposed cysteine of thioredoxins than for glutaredoxins (*68, 69*). Our thiol titration result is in close agreement with the published value of an orthologous protein (Mj0307) from *Methanococcus jannaschii* (pKa 6.27) with which MTH0807 has 51% sequence identity (*16*). Our results also agree closely with the value (6.7) obtained for its paralogue - MTH0895. The low pKa value is also consistent with the strong oxidative properties shown by MTH0807 in the thioredoxin activity test (*4, 16*).

Understanding Thioredoxin Activity in MTH0807

The thioredoxin activity test used here allows the two steps involved in the reaction to be monitored separately. Thus one may ascertain which step is taking place and which is not. The first part of the reaction involves the binding and interaction with thioredoxin reductase and our results clearly indicate that this step is taking place for all the four proteins used in the experiment. This observation confirms the thioredoxin-like nature of MTH0807 and the extent of the oxidation seen here (which is quite high) is in close agreement with what would be expected on the basis of results recently published by Mossner et al. (*4*). These workers demonstrated that when the dipeptide sequence between the active site cysteines in a thioredoxin is mutated from the wild type (Cys-Gly-Pro-Cys) to those of a glutaredoxin (Cys-Pro-Tyr-Cys), a protein disulfide isomerase (Cys-Gly-His-Cys), or a DsbA (Cys-Pro-His-Cys), all the variants proved to be stronger oxidants than the wild type (*4*). The order of oxidizing strength was reported to be glutaredoxin-type > DsbA-type > PDI-type > wild-type with the most oxidizing mutant having the lowest pKa value for the exposed active site cysteine (pKa of 5.9). A similar result (*70*) was obtained when the Pro in the dipeptide sequence between the active site cysteines in *E. coli* was mutated to His to mimic that of PDI (protein disulfide isomerase). The mutagenised thioredoxin had an increased redox potential resulting in significant increase in ability to serve as a substrate for thioredoxin reductase and a decrease in its ability to reduce protein disulfides (*70*). The active site sequence of MTH0807 is the same as that of a glutaredoxin (Cys-Pro-Tyr-Cys), so it is not surprising that MTH0807 exhibits strong oxidizing activity. The second step in the reaction, a reduction step, is almost non-existent for MTH0807 as compared to the control proteins. This is expected for a molecule which is such a strong oxidant. This result may also explain why MTH0807 from *M. thermoautrophicum* (Strain Marburg) was originally reported as not being a substrate for thioredoxin reductase (*1*). In all the thioredoxin/thioredoxin reductase assays used by McFarlan and co-workers, (*1*) it was the reduction step, not the binding step, which was measured.

Understanding Glutaredoxin Activity in MTH0807

To more fully understand the apparent glutaredoxin activity in MTH0807, it is perhaps worthwhile reviewing what is known about a much better characterized glutaredoxin – *E. coli* glutaredoxin. *E. coli* glutaredoxin activity is attributed to residues 12 -13 in the active site (Pro-Tyr), 59 - 60 (Val – Pro) and 69 – 71(Gly - Gly -Tyr), which form a hydrophobic surface on one side of the redox-active site. This hydrophobic surface has been suggested to be the binding site for glutathione and other proteins for the redox reaction (*48*). It is also postulated that Arg8, Lys18 and Asp19 in *E. coli* glutaredoxin, which are located at the opposite sides of the redox active disulphide bridge, may be involved in the actual redox process. An Asp residue located at the N-terminus of the third helix in *E. coli* glutaredoxin is also found in most glutaredoxins (*17*). These charged residues, which are all near the active site, have been implicated in the specific participation of ionic interactions with glutathione (*71, 72*). The structure of MTH0807 shows similar positions of equivalent residues which may form the same hydrophobic binding surface (residues 15 -16 of the active site (Pro-Tyr), 56-57 (Val-Pro), 67– 69 (Phe – Val – Gly)). However, MTH0807 lacks the positively charged residues at the opposite sides of the redox-active disulfide bridge which are thought to be essential in the actual redox process. In our assays, MTH0807 exhibited some oxidation of glutathione; however this was small compared to that of *E. coli* glutaredoxin and T4 glutaredoxin. The low level of glutathione oxidation which was observed could be attributed to non-specific binding of glutathione and the intrinsic oxidative ability of MTH0807. Glutathione is essential for glutaredoxin function and a literature search revealed that most archaeabacteria, including *Methanobacterium thermoautotrophicum* do not contain glutathione or glutathione-like cystolic thiols (*1, 16, 17*). Since glutathione has not been isolated in this organism, it seems reasonable to suggest that the activity obtained was perhaps more an artifact of its structural or sequential similarity to glutaredoxin than its evolved function. However, it could also be argued that this activity is an early precursor towards glutaredoxin activity.

Functional relationship between MTH0807 and MTH0895.

An important question that may be asked is why does a simple archaeon like *M. thermoautotrophicum* have two thioredoxins? It is important to note that most eubacteria and all eukaryotes have both a thioredoxin and a glutaredoxin encoded in their genomes. This redundancy in redox function for these organisms may have arisen because the activities sustained by glutaredoxin and thioredoxin (DNA synthesis, transcriptional control and protein function regulation) are so vital to an organism's viability. The same may be true for *M. thermoautotrophicum* and other archeons too.

Structural characterization and biochemical assay comparisons between MTH0807 and MTH0895 indicate that both proteins have similar structures and both use the same reducing system (the thioredoxin system). MTH0807 is, however, more oxidative and seem to have very little reductive ability as compared to MTH0895. Evidence suggests that MTH0807 is the much more abundant protein in *M. thermoautotrophicum*, especially when grown under native-like anaerobic conditions (*1*). The strong reducing environment in which this organism thrives (*16*) suggests the need for a strongly oxidizing form of thioredoxin to provide the reducing equivalents to ribonucleotide reductase for DNA synthesis. It appears that MTH0807 fits the bill. In contrast, MTH0895 which is the less abundant protein may be a back-up to MTH0807 or it may be involved in other disulfide redox activities within the cytoplasm. Alternately, MTH0895 may be the thioredoxin that is produced when *M. thermoautotrophicum* is exposed to more aerobic conditions.

While both MTH0895 and MTH0807 are true thioredoxins, in some respects, MTH0807 seems to play the role of a glutaredoxin (with its strong oxidizing potential and other physical features) while MTH0895 seems to play the role of a thioredoxin. Given the "archaic" nature of Archea and their hypothesized role in eubacterial and eukaryotic evolution, we would suggest that MTH0807 and MTH0895 may be molecular "fossils" representing an early branch point in the evolution of glutaredoxins and thioredoxins. One could speculate that MTH0807 (or a related ancient variant) maintained its size, shape and active site residues but changed certain adjacent charged residues to become a glutaredoxin, while MTH0895 (or a related ancient variant) lengthened and slightly changed its active site to become a more effective thioredoxin. The dual characteristics (thioredoxin-like and glutaredoxin-like) shown by MTH0807, such as the sequence and active site similarities to glutaredoxins (*1*), overall structure similarity to glutaredoxins, but with structural details and enzymology more akin to thioredoxins are all factors which would seem to support this assertion. Glutathione (GSH) is known to be absent in *M. thermoautotrophicum* (*1, 16, 17*) however, it is widely distributed in high concentrations (0.1 mM-10 mM) in animal tissues, plants and microorganisms (*73, 74*). It might be that when glutathione became more abundant 1-2 billion years ago, MTH0807-like proteins may have evolved towards a glutaredoxin-like function, using GSH to provide the reducing equivalents for ribonucleotide reductase for DNA synthesis. The relatively small number of residue changes (positively charged residues around the active site) that may be needed to convert MTH0807 to a fully functional glutaredoxin (*1*) suggests that this evolutionary process may have been relatively quick and simple.

3.5 Conclusion

This project was undertaken to clarify/identify the possible redox role of MTH0807. Earlier workers had found conflicting evidence about the possible functions for this protein and had suggested that MTH0807 may be part of a distinct ribonucleotide-reducing system involving ferredoxins and ferredoxin-thioredoxin reductases (*1*). This suggestion arose because of the apparent inactivity between MTH0807 and thioredoxin reductase. Using an improved, two-part thioredoxin reductase assay we have shown that MTH0807 does indeed interact with thioredoxin reductase. In addition, functional and structural comparisons of MTH0807 to an orthologous (51% identity) thioredoxin from *M. jannaschi* as well as detailed analysis of its own surface topology indicate that MTH0807 is indeed a true thioredoxin and that it belongs to the thioredoxin-thioredoxin reductase system of redox proteins. To further confirm this assessment, we have conducted sequence searches against the *M. thermoautotrophicum* genome to attempt to identify possible ferredoxin-thioredoxin homologues and have found no significant (E<0.0 1) hits. This evidence strongly indicates that MTH0807 is not part of a distinct ribonucleotide-reducing system as had been postulated earlier. However, the unusual activity, sequence and structure of MTH0807 (and its paralogue MTH0895) have led us to speculate that these two molecules may represent a group of ancient proteins that were ancestral to both thioredoxins and glutaredoxins.

3.6 References

1. McFarlan, S. C., Terrell, C. A. and Hogenkamp, P. C. (1992) *J. Biol. Chem.*, **267**, 10561 – 10569.
2. Martin, J. L. (1995) *Structure*, **3**, 245 – 250.
3. Holmgren, A. (1995) *Structure*, **3**, 239 – 243.
4. Mossner, E., Huber-Wunderlich, M. and Glockshuber, R. (1998) *Protein Science*, **7**, 1233 – 1244.
5. Kallis, G. B. and Holmgren, A. (1980) *J. Biol. Chem.*, **255**, 10261 – 10265.
6. Holmgren, A. (1989) *J. Biol. Chem.*, **264**, 13963 – 13966.
7. Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R. Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D. and Reeve, J.N. (1997) *J. Bacteriol.*, **179**, 7135 – 7155.
8. Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K. and Kikuchi, H. (1998) *DNA Res.*, **5**, 147 – 177.
9. Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C., Fleischmann, R.D., Quack-enbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirkness, E.F., Dougherty, B.A., Mckenny, K., Adams, M.D., Loftus, B. and Venter, J.C. (1997) *Nature*, **392**, 353 – 358.

10. Kawarabayasi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y., Jin-no, K., Taakahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Nakazawa, H., Takamiya, M., Masuda, S., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A. and Kikuchi, H. (1999) *DNA Res.*, **6**, 83 – 101.

11. Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D. E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olsen, G.J. and Swanson, R.V. (1998) *Nature*, **392**, 353 – 358.

12. Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heilderberg, J., Sutton, G.G., Fleischmann, R.D., Eisen, J.A. and Fraser, C.M. (1999) *Nature*, **399**, 323 – 329.

13. Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., Fitzgerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Doughgherty, B.A., Tomb, J.F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L, Geoghagen, N.S.M. And Venter, J.C. (1996) *Science*, **273**, 1058 – 1073.

14. Jordan, A., Pontis E., Aslund, F., Hellman, U., Gigert, I. and Reichard, P. (1996) *J. Biol. Chem.*, **271**, 8779 – 8785.

15. Jordan, A., Aslund, F., Pontis, E., Reichard, P. and Holmgren, A. (1997) *J. Biol. Chem.*, **272**, 18044 – 18050.

16. Lee, D. Y., Ahn, B-Y. and Kim, K-S. (2000) *Biochemistry*, **39**, 6652 – 6659.

17. Bhattacharyya, S., Habibi-Nazhad, B., Amegbey, G., Slupsky, C.M., Yee, A., Arrowsmith, C. and Wishart D.S. (2002) *Biochemistry*, **41**, 4760 – 4770.

18. Stehr, M., Schneider, G., Aslund, F., Holmgren, A. and Lindquist, Y. (2001) *J. Biol. Chem.*, **276**, 35836 – 35841.

19. The QIAexpressionist™, A handbook for high-level expression and purification of 6x His-tagged proteins, Fifth Edition March 2001 pp 63-73 QIAGEN®

20. Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J.R., Booth, V., Mackereth, C.D., Saridakis, V., Ekiel, I.., Kozlov, G., Maxwell, K.L., Wu. N., McIntosh, P., Gehring, K., Kennedy, M.A., Davidson, A.R., Pai, F.E., Gerstein, M., Edwards, A.M. and Arrowsmith, C.H. (2000) *Nat. Struct. Biol.*, **7**, 903 – 909.

21. Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Le, B., Ramelot, T., Lee, G.M., Bhattacharyya, S., Gutierrez, P., Denisov, A., Lee, C-H., Cort, J.R., Kozlov, G., Liao, J., Finak, G., Chen, L., Wishart, D., Lee, W., McIntosh, L.P., Gehring, K., Kennedy, M.A., Edwards, A.M. and Arrowsmith, C.H. (2002) *Proc. Nat. Acad. Sci. USA*, **99**, 1825 – 1830.

22. Jeener, J., Meier, B.H., Bachmann, P. and Ernst, R.R. (1979) *J. Chem. Phys.*, **71**, 4546 – 4553.

23. Kay, L.E., Keifer, P. and Saarinen, T. (1992) *J. Am. Chem. Soc.*, **114**, 10663 – 10665.

24. Zhang, O., Kay, L.E., Olivier, J.P. and Forman-Kay, J.D. (1994) *J. Biomol. NMR*, **4**, 845 – 858.

25. Kuboniwa, H., Grzesiek, S., Delagio, F. and Bax, A. (1994) *J. Biomol. NMR*, **4**, 871 – 878.

26. Kay, L.E., Xu, G.Y. and Yamazaki, T. (1994) *J. Magn. Reson.*, **109**, 129 – 133.

27. Wishart, D.S. and Case, D.A., (2001) *Meth. Enzymol.*, **338**, 3 – 34.

28. Wishart, D.S. and Sykes, B.D. (1994) *Meth. Enzymol.*, **239**, 363 – 392.

29. Brunger, A.T. (1993) X-PLOR manual version 3.851, Yale University, New Haven, CT.

30. Nilges, M. Gronenborn, A.M., Brunger, A.T. and Clore, G.M. (1988) *Protein Eng.*, **2**, 27 – 38.

31. Nigles, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.*, **229**, 317 – 324.

32. Wuthrich, K. (1986) NMR of Proteins and Nucleic Acids, John Wiley and Sons Inc. New York

33. Nilges, M., Kuszewski, J. and Brunger, A.T. (1991) In: Computational Aspects of the study of Biological Macromolecules by NMR, (J.C. Hoch, ed.) New York: Plenum Press.

34. Kuszewski, J., Nilges, M. and Brunger, A.T. (1992) *J. Biomol. NMR*, **2**, 33 – 56.

35. Garrett, Kuszewski, J., Hancock, Lodi, Vuister, Gronenborn, A.M. and Clore, G.M. (1994) *J. Mag. Reson.*, **104**, 99 – 103.

36. Kuszewski, J., Qin, Gronenborn, A. M. and Clore, G. M. (1995) *J. Mag. Reson.*, **106**, 92 – 106.

37. Kuszewski, J., Gronenborn, A.M. and Clore, G.M. (1995) *J. Mag. Reson.*, **107**, 293 – 297.

38. Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R. and Thornton, J. (1996) *J. Biomol. NMR*, **8**, 477 – 486.

39. Willard, L., Anuj, R., Zhang, H., Monzavi, H., Boyko, R., Sykes, B.D. and

Wishart, D.S. (2003) *Nucl. Acids Res.*, **31**, 3316 – 3319.

40. Koradi, R., Billeter, M. and Wuthrich, K. (1996) *J. Mol. Graphics*, **14**, 51 – 55.
41. Dyson, J.H., Jeng, M-F, Tennant, I.S., Lindell, M., Cui, D-S., Kuprin, S. and Holmgren, A. (1997) *Biochemistry*, **30**, 2622 – 2636.
42. Holmgren, A. (1979) *J. Biol. Chem.*, **254**, 3664 – 3671.
43. Wang, Y. Ph.D. Thesis (1999) University of Alberta, Edmonton, Alberta.
44. Cave, W.J., Cho, S.H., Batchelder, M.A., Yokota, H., Kim, R. and Wemmer, E.D. (2001) *Prot. Sci.*, **10**, 384 – 396.
45. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) *Nuc. Acid Res.*, **25**, 3389 – 3402.
46. Shindyalov, I.N. and Bourne, P.E. (1998) *Protein Eng.*, **11**, 739 – 747.
47. Lennon, B.W., Williams Jr., C. H. and Ludwig, M.L. (2000) *Sci.*, **289**, 1190 – 1194.
48. Eklund, H., Cambillau, C., Sjoberg, B.M., Holmgren, A., Jornvall, H., Hoog, J.O. and Branden, C.I. (1984) *EMBO J.*, **3**, 143 – 1449.
49. Holmgren, A., Kallis, G.B. and Nordstrom, B. (1981) *J. Biol. Chem.*, **256**, 3118 – 3124.
50. Russel, M. and Model, P. (1986) *J. Biol. Chem.*, **261**, 14997 – 15005.
51. Holm, L. and Sander, C. (1993) *J. Mol. Biol.*, **233**, 123 – 138.
52. Holm, L. and Sander, C. (1996) *Science*, **273**, 595 – 602.
53. Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) *J. Mol. Biol.*, **247**, 536 – 540.
54. Polgar, L. (1974) *FEBS Lett.*, **38**, 189 – 190.
55. Nelson, J.W. and Creighton, T.E. (1994) *Biochemistry*, **33**, 5974 – 5983.

56. Benesch, R.E. and Benesch, R. (1995) *J. Am. Chem. Soc.*, **77**, 5877 – 5881.

57. Holmgren, A. and Bjornstedt, M. (1995) *Meth. Enzymol.*, **252**, 1999 – 2008.

58. Waksman, G., Krishna, T.S.R., Williams, C.H. and Kuriyan, J. (1994) *J. Mol. Biol.*, **236**, 800.

59. Holmgren, A. (1978) *J. Biol. Chem.*, **253**, 7424 – 7430.

60. Szajewski, R.P. and Whitesides, G.M. (1980) *J. Am. Chem. Soc.*, **102**, 2011 – 2026.

61. Holmgren, A. (1979) *J. Biol. Chem.*, **254**, 3672 – 3678.

62. Takahashi, N. and Creighton, T.E. (1996) *Biochemistry*, **35**, 8342 – 8353.

63. Gan, Z.R., Polokoff, M.A., Jacobs, J.W. and Sardana, M.K. (1990) *Biochem. Biophys. Res. Comm.*, **168**, 944 – 951.

64. Mieyal, J.J., Starke, D.W., Gravina, S.A. and Hocevar, B.A. (1991) *Biochemistry*, **30**, 8883 – 8891.

65. Vohnik, S., Hanson, C., Tuma, R., Fuchs, J.A., Woodward, C. and Thomas, G.J. Jr. (1998) *Prot. Sci.*, **7**, 193 – 200.

66. Wilson, N.A., Barbar, E., Fuchs, J.A. and Woodward, C. (1995) *Biochemistry*, **34**, 8931 – 8939.

67. Jeng, M.F., Holmgren, A. and Dyson, H.J. (1995) *Biochemistry*, **34**, 10101 – 10105.

68. Gan, Z.R. and Wells, W.W. (1987) *J. Biol. Chem.*, **262**, 6704 – 6707.

69. Sun, C.H., Berardi, M.J. and Bushweller, J.H. (1998) *J. mol. Biol.*, **280**, 687 – 701.

70. Krause, G., Lundstrum, J., Barea, J.L., Pueyo de la Cuesta, C. and Holmgren, A. (1991) *J. Biol. Chem.*, **266**, 9494 – 9500.

71. Nordstrand, K., Aslund, F., Holmgren, A., Otting, G., and Berndt, K.D. (1999) *J. Mol.Biol.*, **286**, 541 – 552.

72. Berardi, J.M. and Bushweller, J.H. (1999) *J. Mol. Biol.*, **292**, 151 – 161.

73. Meister, A. (1988) *J. Biol. Chem.*, **263**, 17205 – 17208.

74. Fuchs, J. A. (1989) In: Glutathione: Chemical, Biochemical and Medical Aspects, (Dolphin, D., Poulsen, R., and Avramovic, O. eds.) Part B pp 551 – 570, John Wiley and Sons, New York.

Table 3.1 Structural statistics for MTH0807: (final 20 models).

Total Restraints used	997
Total NOE Restraints	863
Intra-residue	214
Sequential (|i-j| = 1)	345
Medium range (1 < |i-j| < 5)	187
Long range (|i-j| > 4)	127
# Hydrogen bonds	42
# Dihedral Restraints	82
Coordinate precision (Angstroms)	
RMSD of all backbone atoms	0.55 ± 0.21 Å
RMSD of all heavy atoms	1.09 ± 0.33 Å
Procheck NMR statistics	
Residues in most favored region	67.0%
Residues in additional allowed region	30.6%
Residues in generously allowed region	1.3%
Residues in disallowed region	1.1%

Table 3.2 Protein structures similar to MTH0807 derived from the SCOP database:

Protein	**Z Score**	**RMSD**	**Aligned length**	**Seq. length**	**No of gaps**	**PDB code**
Thioredoxin MTH0895 (*Methanobacterium thermoautotrophicum*)	8.58	2.90	60	77	5	1ilo
Thioredoxin Mj0307 (*Methanococcus jannaschii*	8.47	3.20	69	85	6	1fo5
Glutaredoxin (*Escherichia coli*)	8.40	3.30	64	85	6	1egr
Glutaredoxin (Bacteriophage T4)	8.31	3.30	64	87	7	1de2
Thioredoxin H (*Chlamydomonas reinhardt*	8.00	3.50	74	112	4	1tof
Thioredoxin-like(2Fe-2S) Ferredoxin (*Aquifex Aeolicus*)	7.80	3.20	64	109	6	1f37
Glutaredoxin (Bacteriophage T4)	7.60	3.40	65	87	6	1aaz
Glutathione transferase P1- (*Homo sapiens*)	7.50	3.70	67	209	5	1eoh
Glutathione transferase P1- Complexed with cibacron Blue (*Homo sapiens*)	7.40	3.70	66	208	6	20gsb
Thioredoxin (*Homo sapiens*)	7.40	3.30	71	105	5	3trx

Figure 3.1 pET-15b cloning/expression system from Novagen. The MTH0807 gene was sub-cloned into the multiple cloning sites such that the protein is expressed with an additional 20 amino acids at the N-terminus.

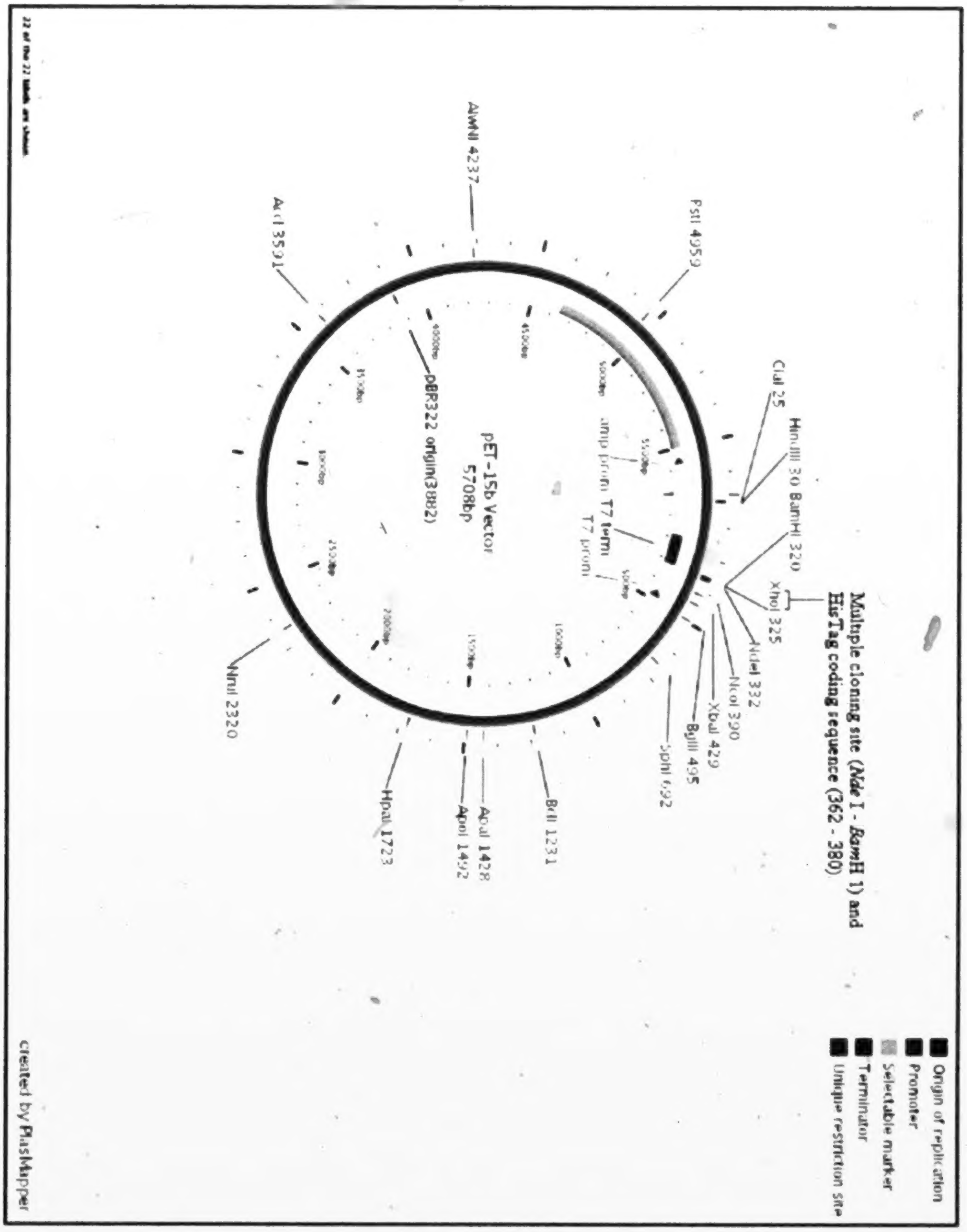

pET-15b Vector
5708bp
ClaI 25
HindIII 30 BamHI 320
XhoI 325
NdeI 332
NcoI 390
XbaI 429
BglII 495
SphI 692
BclI 1231
AvaI 1428
ApoI 1492
HpaI 1723
MluI 2320
AccI 3591
AlwNI 4237
PstI 4959
pBR322 origin(3882)
amp prom
T7 term
T7 prom
Multiple cloning site (*Nde I - BamH 1*) and HisTag coding sequence (362 - 380)
Origin of replication
Promoter
Selectable marker
Terminator
Unique restriction site
22 of the 22 labels are shown
created by PlasMapper

Figure 3.2 ^1H-^{15}N HSQC spectrum from 1 mM MTH0807 in 500 μL of a buffer solution made up of 50 mM NaH_2PO_4, 100 mM NaCl, 50 μL D_2O, 1 mM DSS pH 6.0 and 10 μL of a 3% solution of sodium azide, collected on Varian Inova 500 MHz spectrometer.

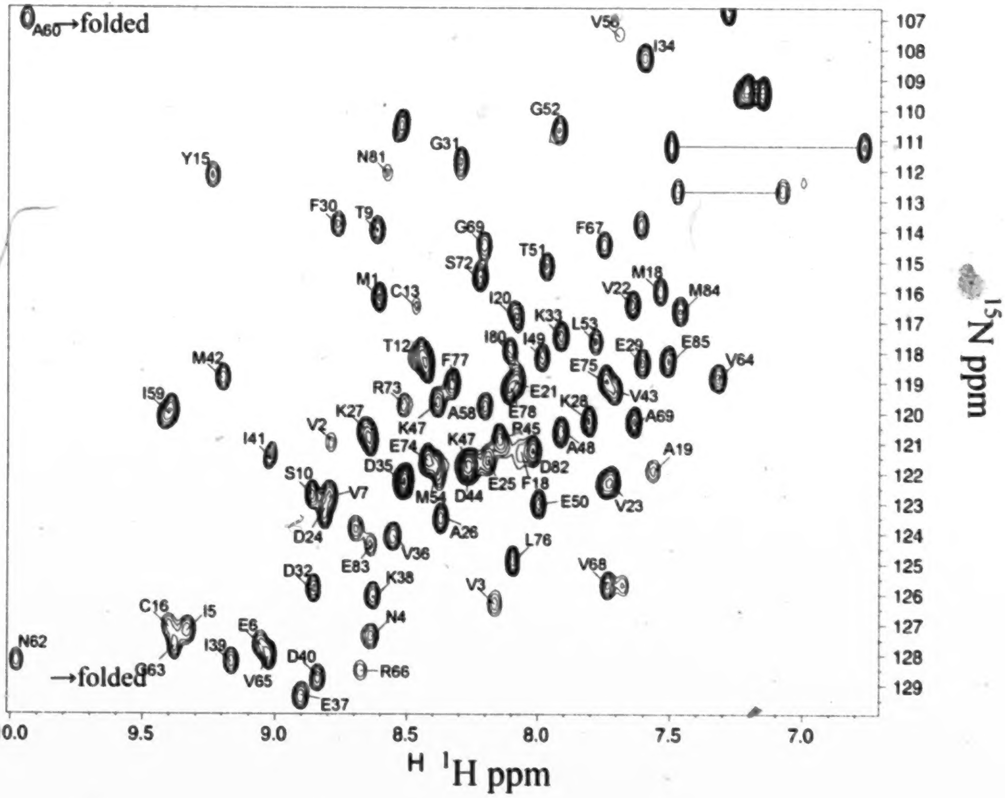


Figure 3.3 A stereoview of the MOLMOL superposition of backbone atoms from the 20 best calculated structures.

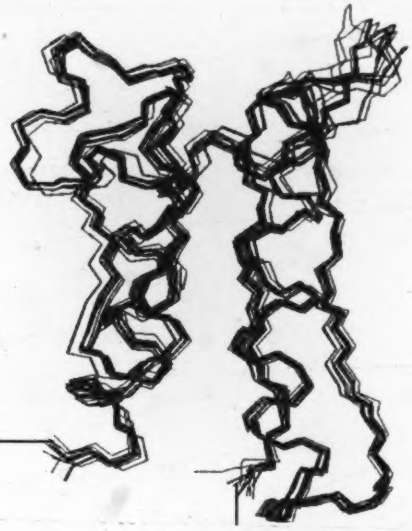

C-Terminus
N-Terminus

Figure 3.4 Ribbon diagram of a representative structure of MTH0807.

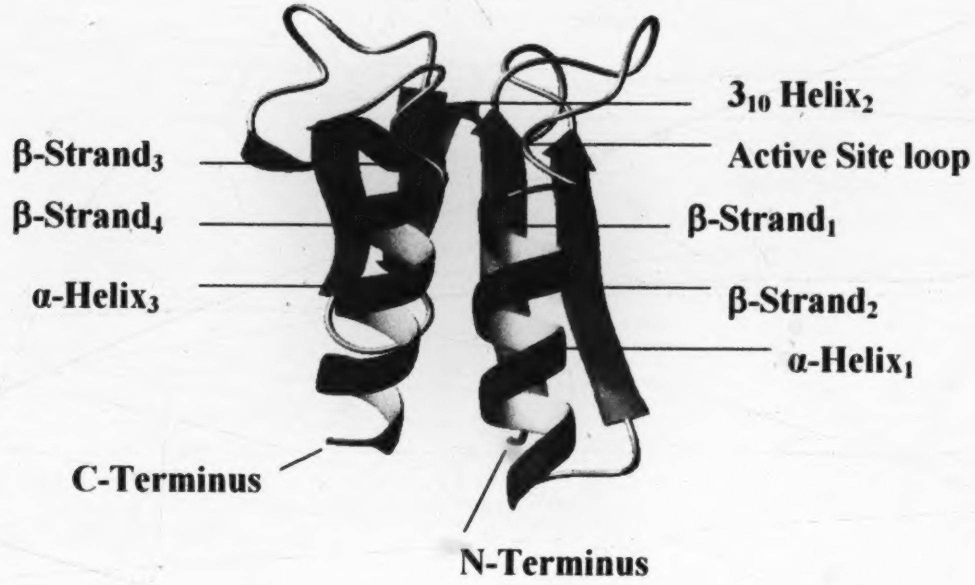

3_{10} Helix$_2$
β-Strand$_3$
Active Site loop
β-Strand$_4$
β-Strand$_1$
α-Helix$_3$
β-Strand$_2$
α-Helix$_1$
C-Terminus
N-Terminus

Figure 3.5 Structure-based sequence alignment obtained from the CE database for MTH0807, thioredoxin H from *Chlamydomonas reinhardtii* (1TOF), thioredoxin from *E. coli* (1XOB), human thioredoxin (4TRX), thioredoxin from *M. thermoautotrophicum* (1ILO), thioredoxin from *M. jannaschii* (1FO5), glutaredoxin from *E. coli* (1EGR) and human glutaredoxin (1JHB). The β-strands and α-helices are marked grey. The 100% conserved residues are enclosed in a box.

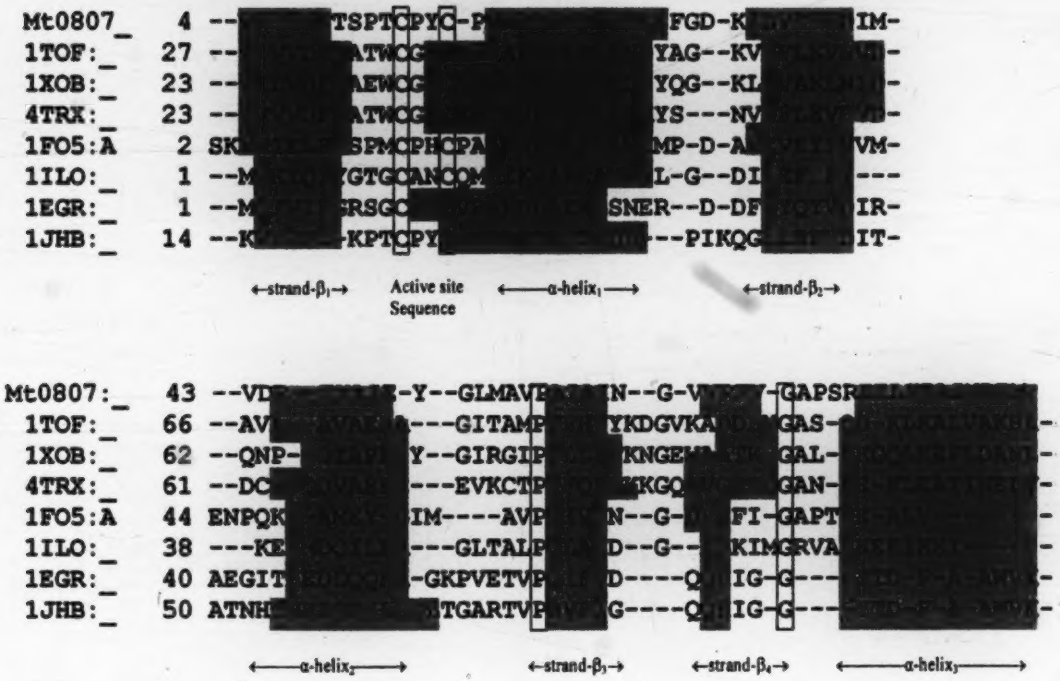

Mt0807_ 4
1TOF:_ 27
1XOB:_ 23
4TRX:_ 23
1FO5:A 2
1ILO:_ 1
1EGR:_ 1
1JHB:_ 14
←strand-β1→
Active site Sequence
←α-helix1→
←strand-β2→
Mt0807:_ 43
1TOF:_ 66
1XOB:_ 62
4TRX:_ 61
1FO5:A 44
1ILO:_ 38
1EGR:_ 40
1JHB:_ 50
←α-helix2→
←strand-β3→
←strand-β4→
←α-helix3→

Figure 3.6 Measurement of thiol ionization in MTH0807 by ultraviolet absorbance at 240 nm. The ε_{240} (extinction coefficient at 240 nm) from pH 5.0 to 10.0 of a solution of MTH0807 containing 30 μM protein was recorded in 0.1 mM EDTA and 100 mM potassium phosphate buffer.

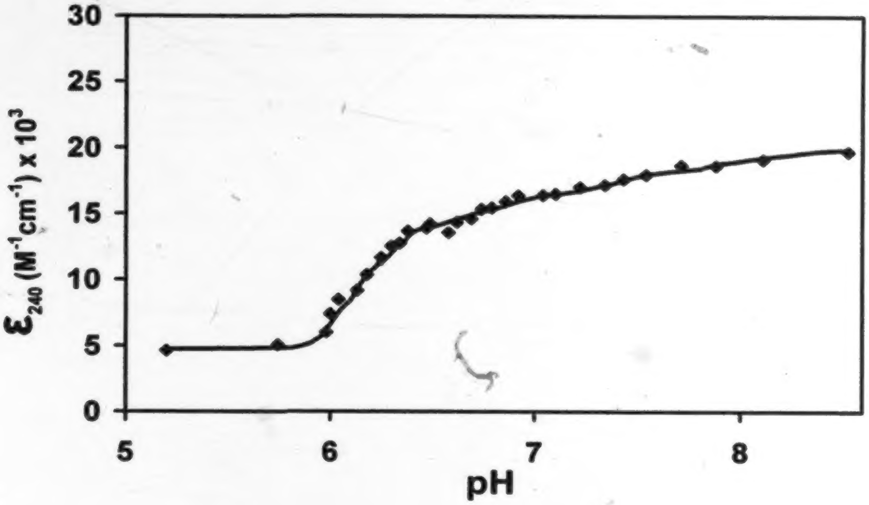
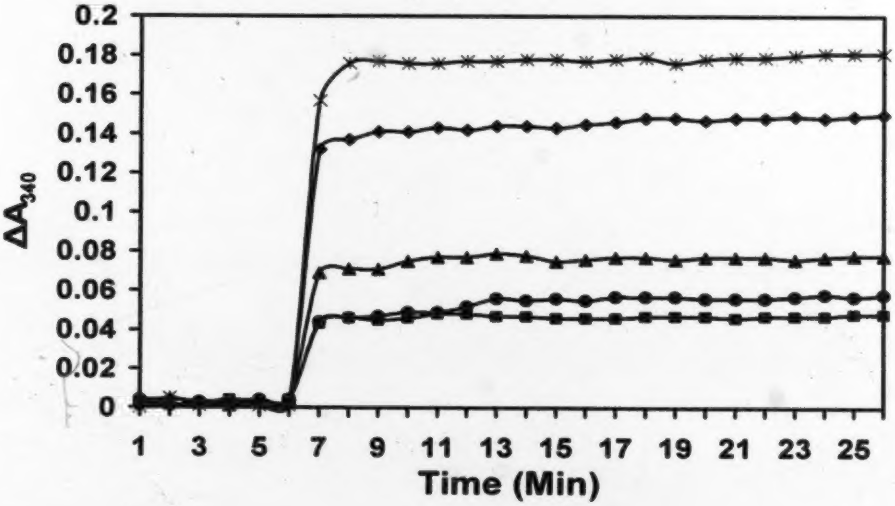
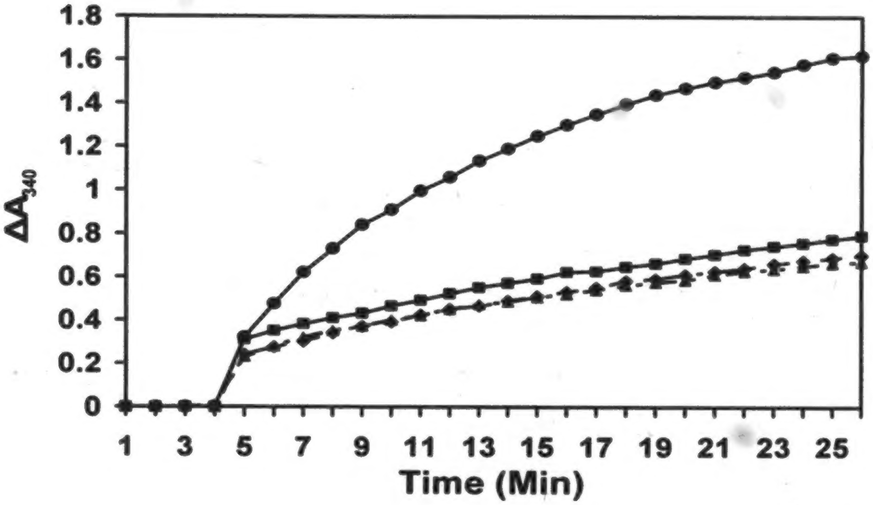


Figure 3.7 Measurement of thioredoxin and glutaredoxin activity via UV absorbance

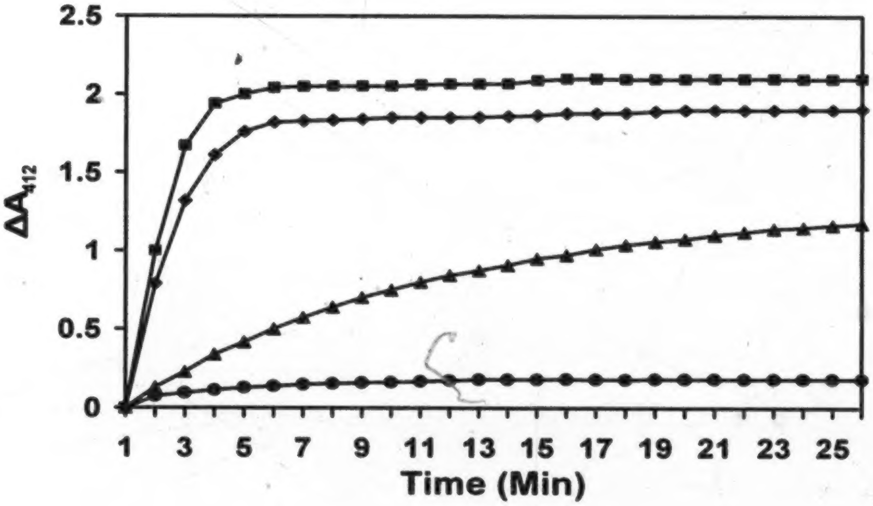
(a) Glutaredoxin activity kinetics. The ΔA_{340} (change in absorbance at 340 nm) for T4 glutaredoxin (✱), *E. coli* glutaredoxin (♦), *E. coli* thioredoxin (■) MTH0895 (•) and MTH0807 (▲).The change in absorbance is proportional to the amount of glutathione oxidised hence a measure of the glutaredoxin activity



(b) Thioredoxin activity kinetics (oxidation of NADPH). The ΔA_{340} (change in absorbance at 340 nm) for T4 glutaredoxin (■), E. coli thioredoxin (♦), MTH0895 (▲) and MTH0807 (•). The change in absorbance is proportional to the amount of NADPH oxidised hence a measure of the extent of thioredoxin/thioredoxin reductase interaction



(c) Thioredoxin activity kinetics (reduction of DTNB). The ΔA_{340} (change in absorbance at 340 nm) for T4 glutaredoxin (■) E. coli thioredoxin (♦), MTH0895 (▲), and MTH0807 (•).The change in absorbance is proportional to the amount of DTNB reduced hence a measure of the reductive ability of the protein



Chapter 4

Structural and Functional Characterisation of MTH0776

4.1 Introduction

Methanobacterium thermoautotrophicum is a methanogenic archaeon that grows optimally at ~65°C. Originally isolated from sewage ponds in southern Illinois, *M. thermoautotrophicum* is a microorganism that biodegrades waste materials and produces natural gas in the process. Due to its potential economic and environmental importance the *M. thermoautotrophicum* genome was sequenced in the mid 1990's (*1*) to better understand the underlying biochemistry and microbiology of its unique biodegradation capabilities. The genome (1,751,377 bp) consists of 1855 open reading frames, with roughly ⅔ of the proteins having assigned functions. Because of the favorable solution properties of proteins isolated from thermophilic bacteria, *M. thermoautotrophicum* has become one of the most actively studied organisms in structural genomics (*2*).

MTH0776 is a 101 amino acid residue protein originally isolated from *M. thermoautotrophicum*. It has been successfully cloned and expressed in *E. coli*. This small protein belongs to a distinct orthologous group (COG 4033) consisting of 11 proteins varying in size from 92-124 residues found only in methanogenic archaebacteria. From extensive comparative genomic analysis, we believe that MTH0776 is part of a two-protein operon with its upstream partner -- MTH0777 (157 amino acid residues), which is also specific to methanogenic archaebacteria. The fact that this two-component operon is conserved in all methane metabolizing archaea, strongly suggests that both proteins have an important role in methanogenesis and that the proteins may act in a similar portion of the pathway, potentially as binding partners. The crystal structure of

MTH0777 was recently determined (PDB accession: 1KJN) and it exhibits an α/β fold to which a function has not yet been assigned. In an effort to better understand its structure, its role in methanogenesis and the molecular interactions between MTH0776 and MTH0777, we have undertaken the NMR structure determination of MTH0776. The structure of MTH0776 represents a novel fold, and our structural and bioinformatics studies show that MTH0776 and MTH0777 may be part of a transport system conserved in the methanogenic archarbacteria.

4.2 Materials and Methods

Protein Expression, Purification and Sample Preparation

The MTH0776 gene was sub-cloned into a pET15b vector (Novagen) and transformed into BL21 (DE3) *E. coli* cells for expression. The vector expresses the protein with an N-terminal His_6 tag followed by a thrombin cleavage site. This adds an additional 20 amino acids to the N-terminus of the native sequence of the protein (figure 4.01). Protein expression was accomplished by growing the transformed *E. coli* cells in a minimal (M9) growth medium. Uniform isotopic enrichment with ^{15}N or $^{15}N/^{13}C$ was accomplished by substituting the NH_4Cl and/or glucose in standard M9 media with $^{15}NH_4Cl$ and/or [U-^{13}C]-glucose (Martek Labs, Richmond VA) respectively. Cells were grown at 37 °C in shaker flasks to an optical density of 1.0 and then induced with isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. The cells were allowed to grow for an additional eight hours before harvesting and centrifugation. To release the protein, the cell pellets were re-suspended in a lysis buffer (100 mM NaH_2PO_4, 10 mM Tris.Cl, 8 M Urea, pH 8.0). The volume of lysis buffer added was 5 mL per gram (wet weight) of cell pellet. The mixture was then stirred for 45 minutes at room temperature, until the solution became translucent, signifying complete lysis. The lysed cells were then centrifuged at 15,000 rpm for 30 minutes at room temperature to pellet any cellular debris. The supernatant was isolated and the protein was purified and re-natured using a nickel nitrilotriacetic acid-agarose (Ni-NTA) column (Qiagen) as described elsewhere (*3*). Specifically, the supernatant (lysate) was passed over a nickel nitrilotriacetic acid-agarose (Ni-NTA) column (Qiagen) and washed with the lysis buffer. With the protein still attached to the column, it was washed with wash buffer (100 mM NaH_2PO_4, 10 mM Tris.Cl, 8 M Urea, pH 6.3). The column was then washed with 20 ml of a buffer solution made up of (500 mM NaCL, 20 mM Tris.Cl, 10 mM β-mercaptoethanol, 20% glycerol, 20 mM Imidazole and 8 M Urea, pH 8.0). Refolding of the bound protein was then performed by the use of a linear 8 – 0 M urea gradient, starting with the wash buffer above (500 mM NaCL, 20 mM Tris.Cl, 10 mM β-mercaptoethanol, 20% glycerol, 20 mM Imidazole and 8 M Urea, pH 8.0) and ending with one without urea (500 mM NaCL, 20 mM Tris.Cl, 10 mM β-mercaptoethanol, 20% glycerol and 20 mM Imidazole, pH 8.0) over a two hour period. The column was washed with 50 mL buffer (500 mM NaCL, 20 mM Tris.Cl, 10 mM β-mercaptoethanol, 20% glycerol and 20 mM Imidazole, pH 8.0) and the refolded recombinant protein eluted from the column with a buffer consisting of 500 mM NaCL, 25 mM Na_2HPO_4 and 500 mM Imidazole, pH 8.0. DTT was immediately added to the eluted protein to a concentration of 20 mM to prevent the oxidation and precipitation of the protein. The isolation and purification process outlined above yielded approximately 30mg of histidine tagged protein per litre of minimal (M9) growth medium as determined by UV absorbance at 280 nm using an extinction coefficient of 5240 M^{-1} cm^{-1}.

Protein samples for functional assays/enzymatic screening were dialysed into a buffer (pH 7.5) solution consisting of 50 mM HEPES and 300 mM NaCl. NMR samples were prepared by dissolving about 15 mg of the His-tagged protein in 500 μL of a buffer (pH 6.8) made up of 20 mM NaH_2PO_4, 300 mM NaCl, 50 μL of D_2O, 1 mM 2, 2 - dimethyl-2-silapentane-5-sulfonic acid (DSS), 15 mM DTT and 10 μL of a 3% solution of sodium azide. Spectra collected with and without the histidine tag were found to be essentially identical (data not shown), indicating that the poly-His tag does not interfere with the spectral or structural properties of the protein (*4*).

Purified samples of ^{15}N-labelled MTH0776 and unlabelled MTH0777 were prepared as described above and excess MTH0777 was then added to a 1mM solution of MTH0776 and dialysed into a buffer solution consisting of 50 mM NaH_2PO_4, 300 mM NaCl and 15mM DTT. The protein mixture was then prepared for a ^{15}N-HSQC experiment (Figure 4.02) and the results compared to that of a pure 2 mM solution of ^{15}N-labelled MTH0776 protein sample (Figure 4.03). Samples for ^{15}N T_2 measurement were prepared in the same way as above, with the comcentration of MTH0776 in both samples being 0.5 mM and that of MTH0777 being 1.0 mM. Another sample prepared in a similar way as the MTH0776/MTH0777 mixture, was passed through a Biogel™ P-30 size exclusion gel column to ascertain if the putative MTH0776/MTH0777 complex elute from the column together.

NMR Spectroscopy, Chemical shift assignment and Experimantal Restraints

NMR experiments were recorded at 25 °C on Varian Inova 500 and 800 MHz spectrometers equipped with a 5 mm triple resonance and pulse gradient accessories. All spectra were processed with VNMR software or with NMRPIPE (*5*). Proton chemical shifts were referenced to internal DSS while ^{13}C and ^{15}N chemical shifts were referenced indirectly to DSS using the absolute frequency ratios (*6*).

Sequential chemical shift assignments were obtained by identifying ^{13}Cα(i) to ^{13}Cα (i-1) and ^{13}Cβ(i) to ^{13}Cβ(i-1) connectivities from the HNCACB (*7*) spectrum. These were confirmed and any existing ambiguities resolved using HNCA, CBCA(CO)NH spectra (*8*) and NOEs measured from 2D homonuclear ^{1}H-^{1}H NOESY and 3D ^{1}H-^{15}N NOESY-HSQC (*9*) spectra. The backbone chemical shift assignment was then completed using additional data from HNCO and HNHA spectra (*11*). Side chain ^{1}H assignments for each amino acid were then added using data from a DE-HCCH_TOCSY experiment as well as C(CO)NH and H(CCO)NH spectra (*11*). The side chain ^{1}H assignments for the aromatic residues were identified and assigned from the 2D homonuclear ^{1}H-^{1}H NOESY, 2D ^{13}C-HSQC and 3D ^{13}C edited NOESY experiment optimized for the aromatic region. Stereo-specific assignments of ^{1}Hβ protons were based on the intensities of ^{1}HN-^{1}Hβ and ^{1}Hα-^{1}Hβ cross-peaks in the ^{1}H-^{15}N NOESY-HSQC and 2-D ^{1}H-NOESY spectra. The methyl groups of Val and Leu were assigned stereo-specifically based on the intensity of ^{1}HN-^{1}Hγ, ^{1}Hα–^{1}Hγ cross peaks. All ^{1}H, ^{15}N and ^{13}C polypeptide backbone resonances were assigned with the exception of Pro 56 and Pro 101, and the amino acid side chain assignments for non-labile hydrogens are complete except for the Hε atoms of Methionines giving an overall assignment of about

99%. The ^{1}H, ^{13}C and ^{15}N chemical shifts have been deposited in the BioMagRes Bank under the BMRB accession number 6272 and published (*12*).

NOE distance restraints were obtained using 3D ^{15}N-NOESY HSQC, 3D ^{13}C-NOESY HSQC as well as two-dimensional ^{1}H-^{1}H NOESY spectra, all recorded with a mixing time of 80 ms and 110 ms. The assigned NOE restraints were classified into four distance ranges: 1.8-2.7Å, 1.8-3.5 Å, 1.8-5.0 Å, and 1.8-6.0 Å corresponding to strong, medium, weak and very weak NOE intensities, respectively. NOE peak intensities were measured by volume integration of well resolved peaks. Pseudo-atom corrections were added to the upper distance limits where appropriate (*13*). A 0.5 Å correction was applied to the upper bounds for NOEs involving nonstereospecifically assigned methylene protons, 1.0 Å for NOEs involving methyl protons and 2.0 Å for NOEs involving nonstereospecifically assigned protons on opposite sides of aromatic rings. Torsion angle restraints were predicted using TALOS (*14*) based on the observed ^{13}Cα, ^{13}Cβ, ^{1}Hα, and ^{1}HN chemical shifts. A total of 84 ϕ backbone torsion angle restraints were used in combination with the $^3J_{HNH\alpha}$ coupling constants obtained from a 3D HNHA experiment. Backbone ϕ angles were assigned an uncertainty of ± 10° for residues in well-defined helical or beta-sheet regions. A total of 80 backbone ψ dihedral angle restraints were derived via TALOS and assigned an uncertainty of ± 30°. Hydrogen bond restraints ($d_{O\text{-}HN}$ = 1.8-2.7 Å and $d_{O\text{-}N}$ = 2.8-3.7 Å for each hydrogen bond) for slowly exchanging amide protons were identified via several methods including a D_2O exchange ^{15}N HSQC experiment, the pattern of sequential and interstrand NOEs involving ^{1}HN and C_αH protons (*13*) and from the chemical shift index (*15*).

Structure calculation

Structures for MTH0776 were calculated using X-PLOR-NIH 2.9.6 (*16*). Note that the structures presented here include a histidine residue at the N-terminus in addition to the native sequence of the protein. Using the chemical shift index (*15*), regions of secondary structures in the protein sequence were identified from the assigned chemical shift. Only NOE-derived distance constraints involving the Hβ, Hα and HN atoms of these structured regions along with the full set of intra-residue and sequential NOEs were used in the first stage of structure generation. Initially a set of 60 structures was generated using the *ab initio* simulated annealing protocol (sa_new.inp) applying standard default parameters. After this initial step, several ambiguous long-range NOE assignments were clarified by analyzing the resulting structures with good geometrical and energy parameters. Subsequently the same simulated annealing protocol was repeated with additional and/or corrected NOE data sets to generate a new set of 60 structures. NOE violations were again corrected (usually by extending their upper distance limits) after visually reassessing the corresponding NOE spectral intensities. Typically these problematic NOEs were borderline cases falling between the strong/medium, medium/weak or weak/very weak NOE intensity categories. In the third step, dihedral angles and hydrogen bond restraints were added and the *sa_new.inp* protocols were run again to generate another 60 structures. These structures were then refined using the *refine.inp* protocol in X-PLOR. The final set of 20 structures was selected on the basis that no inter-proton distance restraint violation could be greater than 0.5 Å, no angle violation could be greater than 5°, no bond-length violation could be greater than 0.01Å and no bond angle violations from ideality could be greater than 2°.

A total of 1844 NOE-derived distance restraints (578 long range, 422 medium range, 698 sequential and 146 intra-residue noes), 164 dihedral angle restraints, and 40 hydrogen bond restraints were used to generate the final structural ensemble. The final set of 20 structures was analyzed with PROCHECK-NMR (*17*) and MOLMOL (*18*) and SuperPose (*19*) was used to visualize all the structures and to calculate the RMSD values. The three dimensional coordinates for the final 20 structures have been deposited in the Protein Data Bank (PDB) under accession number 1Z9V and accepted for publication in the Journal of biomolecular NMR.

4.3 Results

Sequence analysis

Sequence analysis of MTH0776 was done with BLAST (*20*), CLUSTALX (*21*) and CGView (*22*). A BLAST search for proteins related to MTH0776 yielded matches which were found only in Archaea. All of these proteins were uncharacterised or hypothetical proteins, belonging to the clustered orthologous group – COG 4033 (Figure 4.04). A multiple sequence alignment of these proteins was done with CLUSTALX and is shown in figure 4.05. Gene position and gene order analysis for MTH0776 and MTH0777 was done using genome maps generated by CGView for a total of 11 different methanogenic organisms and results shows that MTH0776 and MTH0777 are always adjacent to each other although their order does changes (Figures 4.06 – 4.11). A phylogenetic tree (Figure 4.12) was generated from the aligned sequences using CLUSTALX and the neighbour-joining method (*23*) and the results analysed with the boostrap method to provide a robust confidence level for the tree topology. BLAST and

CGView analysis shown that *Methanosarcina acetivorans*, *Methanosarcina mazei*, and *Methanosarcina barkeri* have an extra MTH0776 gene. The phylogenetic tree shows that these three species diverged from a common ancestor after the gene duplication. The conserved proximity of MTH0777 and MTH0776 among all methanogenic archaea suggested that these two proteins may be co-expressed and that they may bind or interact with one another. To check this hypothesis, purified samples of ^{15}N-labelled MTH0776 and unlabelled MTH0777 were prepared. Excess MTH0777 was then added to a 1 mM solution of MTH0776 and dialysed into a buffer consisting of 50 mM NaH_2PO_4, 300 mM NaCl and 15 mM DTT. The protein mixture was then prepared for a ^{15}N-HSQC experiment (Figure 4.02) and the results compared to that of a pure ^{15}N-labelled MTH0776 protein sample in the identical buffer (Figure 4.03). The results shown in Figure 4.13 indicate that MTH0776 and MTH0777 show good evidence of a protein-protein interaction as several residues in the MTH0776 spectrum (L5, E6, T7, L9, I17, A28, H40, I71, T74, K75, C77, Y78 and G79) show marked differences in their chemical shift values and noticeable broadening, relative to other residues. Relaxation (T_2) measurements (appendix 2) also indicate that MTH0776 is in complex with MTH0777 as its average $^{15}NT_2$ relaxation times reduced from 105.08 ± 12.18 miliseconds to 57.40 ± 14.27 miliseconds (a value characteristic of a molecule of about 30 KD in molecular weight). In addition, the sample of MTH0776/MTH0777 mixture eluted from the Biogel P-30 column as a single peak indicating that the two proteins may be in a form of a complex. Of further note, the mixture of MTH0776 and MTH0777 was found to be much more stable than either protein alone in solution. Typically pure MTH0776 solutions will start to precipitate after 3 days, while the pure MTH0777 solutions would last less than 2 days in the same buffer conditions (50 mM NaH_2PO_4, 300 mM NaCl and 15 mM DTT). Conversely the mixture of the two proteins exhibited no precipitation and produced high quality spectra even after several weeks in solution. The majority of the residues being perturbed in the spectrum are nonpolar and solvent exposed (Figure 4.14). Evidently the binding between the two proteins reduces the solvent exposure of these hydrophobic residues, leading to improved solution behaviour.

Enzymatic/Catalytic activity test

In addition to structural, sequential and spectroscopic analysis, samples of purified MTH0776 were tested for catalytic activity using a battery of enzyme assays developed at the University of Toronto. These included phosphatase, esterase/lipase, protease, dehydrogenase, oxidase and phosphodiesterase/nuclease activity. Our results indicated that MTH0776, either alone or in complex with MTH0777 had none of the above activities.

Solution structure

Statistical parameters for the ensemble of 20 calculated structures are presented in Table 4.1. All structures (figure 4.15 and figure 4.16) exhibit good covalent geometry and for all 20 structures, more than 99.0% of the main chain (ϕ, ψ) angles fall in the core or allowed regions of the Ramachandran plot (Table 4.1) as determined using PROCHECK-NMR. MTH0776 is a typical alpha/beta protein and is made up of seven β-strands with three α-helices forming the secondary structural topological arrangement of β_1-α_1-β_2-α_2-β_3-β_4-β_5-β_6-β_7-α_3. The β-strands include residues 2 – 4 (strand β1), 21 – 24 (strand β2), 37 – 43 (strand β3), 47 – 52 (strand β4), 60 – 65 (strand β5), 69 – 74 (strand β6) and 79 – 85 (strand β7). The first helix (α-helix1) is the shortest and runs from residue 9 – 12. The second helix (α-helix2) runs from residue 27 – 33 while the third helix (α-helix3) is the longest and runs from residue 90 – 99. The first β-strand (residues 2 – 4) is however not well defined as it shows up when structure is rendered with RASMOL (*24*) and analysed with SHIFTS (*25*) and vadar (*26*) but does not show up when rendered with MOLMOL and analysed with PROCHECK. The structure of MTH0776 is more or less globular in shape with strands β_1, β_2 and β_4 being antiparallel and on one side while four strands, (β_3 and β_7 (parallel) and β_5 and β_6 (antiparallel) are on the opposite side.

In an attempt to predict the function of MTH0776, a representative structure was submitted to search the CE (*27*), DALI (*28*) and SCOP (*29*) databases. Results obtained from the CE and DALI searches indicate no significant matches to any known structure. SCOP however gave a match with a best Z score of 1.4 and best RMSD of 4.37 to the solution structure of the N-terminal PH/PTB domain of the TFIIH P62 subunit.

4.4 Discussion and Conclusions

The structure of MTH0776 was solved as part of the Structural Proteomics project initiated by the Ontario Cancer Institute in 1999. The primary mandate of this project was to investigate previously unknown, unidentified or unclassified proteins from a thermophilic archeon (*M. thermoautotrophicum*). One of the key objectives of this project has been to identify novel or never-before-seen protein folds. Given the unique ecological niche occupied by this organism and the high percentage of unknown or unclassifiable sequences in its genome (54%, (*1*)), there was a general expectation that many of these "unknown" proteins would yield very novel protein structures. Interestingly, of the ~40 *M. thermoautotrophicum* structures solved to date, less than 10% of the structures exhibit truly novel folds (*30*). MTH0776, therefore, is one of the few proteins from *M. thermoautotrophicum* that exhibits a never-before-seen fold. The discovery of novel protein folds is actually becoming increasingly rare. Statistics collected from the PDB and analyses performed on fold classes have shown a steady decline in the percentage of novel folds deposited into the PDB (http://scop.mrc-lmb.cam.ac.uk/scop/count.html#scop-1.67). On average fewer than 3% of newly deposited proteins have distinct or previously unidentified protein folds.

The identification of novel or unique folds among proteins is important, particularly given the intended role of structural proteomics in addressing the protein folding problem. It is quite apparent that the protein fold universe is finite, with perhaps fewer than 2,000 unique or distinct folds existing among all living organisms (*31*). To date we have found perhaps 1000 of these folds (*32*) and so any extension of the "fold space" provides theoreticians with important new information about what kinds of topologies, packing and secondary structure connections are possible or "allowed".

Of course the other key objective of any structural proteomics initiative is to use structure to help with the determination of function. The prevailing method of inferring the function of a protein from its structure involves using a large number of structural alignment algorithms developed using random sampling methods and similarity measuring functions. Among these algorithms, CE (*27*), DALI (*28*), SCOP (*29*) and VAST (*33*) has been used extensively. However, when a protein has a novel fold, its function can not be inferred based on proteins of known structure. A more recent approach is the use of the protein structure space map (*34*) which was reported to consistently out-perform the other prediction methods. The map was created based on the structure and function of known proteins and proteins sharing similar molecular function were found to colocalize in the protein structure space map. The function of a protein with a novel fold could then be inferred based on the proximity in the map of the protein structure space. It is however in the testing stage and has not been realeased for public use. Given that the structure of MTH0776 appears to be unique, this largely makes it impossible to determine the protein's function via structural homology. Therefore a number of ancillary investigations were undertaken to determine the possible nature/function of MTH0776 through sequence comparisons, binding assays and enzymatic tests.

Disappointingly, a large battery of enzymatic assays covering many common enzymatic functions yielded no useful results, either on MTH0776 alone or in complex with MTH0777. Likewise sequence searches using BLAST, PSI-BLAST and PFAM comparisons yielded little in terms of useful information. Interestingly, the results from our gene order or gene synteny analyses show that both MTH0776 and MTH0777 (or their homologues) were conserved in all methane metabolizing archaebacteria. This level of syntenic conservation is rare and is often a good indication that these two proteins form a two-component operon system. Given that many proteins in two-component operons often physically bind to each other, we chose to investigate the possible interactions between MTH0776 and MTH0777 via NMR. Visual inspection as well as VADAR (*35*) and WHATCHECK (*36*) analysis of the structure of MTH0776 show that majority of the hydrophobic residues are solvent exposed. While this analysis may explain the predicted (http://au.expasy.org/cgi-bin/protparam) and observed instability of MTH0776 in aqueous solution, it could also point to the fact that the molecule is part of a transmembrane complex or part of a multimeric active unit (*36*) and therefore we chose to investigate these possibilities. As can be seen from the superimposed ^{15}N-HSQC spectrum (Figure 4.13) and results from the T_2 measurements (appendix 2), there are substantial changes in the chemical shifts, intensities and linewidths of a number of amino acid residues in the spectra of the ^{15}N-labelled MTH0776 and unlabelled MTH0777 mixture. While more work needs to be done to fully investigate these spectral perturbations, these preliminary data suggests that MTH0776 and MTH0777 form a protein-protein complex in solution.

Since gene analysis, chromatographic studies, structural properties and preliminary NMR studies have shown that MTH0776 and MTH077 bind together, we submitted the X-ray structure of MTH0777 (1KJN) to search the structure database as well. The results show that MTH0777 has a fold similar (Z-Score = 5.9, and RMSD = 2.31) to the KTN domain (K^+ transport, nucleotide binding) (*37*). However MTH0777 lacks the characteristic NADH binding glycine motif (G-X-G-X-X-G-) (*38*) thus explaining the lack of binding of MTH0777 to NADH as we observed in our enzymatic assays. This puts the fold of MTH0777 close to the closely related RCK (regulate the conductance of K^+) domain; however MTH0777 lacks the extra β-α secondary structural element after the 6th alpha helix at the C-terminus of RCK domains (*39*). Both KTN and RCK proteins are cytoplasmic/intracellular domains of transmembrane proteins and contain the Rossman fold (two α-helices (α_1 and α_2) on one side of a six stranded parallel β-sheet and three α-helices (α_3, α_4 and α_5) on the other). The Rossman fold is a very common structural motif found in many different enzymes and ligand binding proteins (*40*). The fold is versatile in their function but have in common the potential to bind small molecules or metal ions and are well known to be involved in uptake or efflux systems that are vital for homeostatic functions critical for cell growth and survival (*41, 42, 43, 44, 45*). The domain is known to be functional by forming quartenary structures usually by hydrophobic association with each other (*39, 40*). In KTN proteins the hydrophobic surfaces for the dimerizations is normally provided by a relatively flat, large (~800Å of buried surface area) hydrophobic patch created by nonpolar residues of the fourth (α_4) and fifth (α_5) helices of the domain (*37*). In RCK proteins this same hydrophobic patch (~1800 Å) is provided by the helix-strand-helix structure (6th α-helix, 7th β-strand and 8th α-helix) attached to the Rossman fold (*39*). It is important to note that mutations of nonpolar amino acid residues to polar ones in these hydrophobic patches, resulted in the production of only insoluble aggregated recombinant proteins of the full length cytoplasmic A subunits of the Ktr K^+ import system of *Methanococcus jannaschii* (Mja218) (*37*). Similar results were obtained in E. coli RCK and BK channel proteins (*39*). A consequence of the dimerization of these domains in solution at such hydrophobic patches would be the elimination of substantial hydrophobic residue exposure to the aqueous environment thus ensuring stability.

It has often been argued that, the three dimensional structures of single macromolecules are often uninformative about function if taken out of context. Just as words must be assembled into sentences, paragraphs, chapters and books to make sense, vital cellular functions are performed by structured ensemble of proteins (protein complexes) not by freely diffusing and occasionally colliding proteins (*46*). If two proteins interact with one another, they usually participate in the same or related cellular function - guilty by association (*47*). This is the principle behind the highly successful yeast two-hybrid screens which has revealed interactions that place functionally unclassified proteins in a biological context, interactions between proteins involved in the same biological function, and interactions that link biological function together into larger cellular processes (*48*).

While a considerable amount of work may need to be done to clearly characterize the nature of the binding between MTH0776 and MTH0777, results so far indicate that binding is taking place. MTH0777 lacks the extra β-α structural element at the end of the 6th α-helix, which is essential in the formation of the quartenary structure necessary for function in RCK domains. Our results show that the majority of the residues being perturbed in the mixture of MTH0776 and MTH0777 are nonpolar residues which could form a surface for which protein-protein interaction could occur between MTH0776 and MTH0777. This may lead to the quartenary structure necessary for function.
A representative structure of MTH0776 submitted to the Profunc and Castp webservers (*49, 50*), revealed the presence of two large clefts (Figure 4.17 and Figure 4.18) on the surface of the protein which are formed primarily by the side chains of residues perturbed in figure 4.14. These clefts (the two largest on the surface of the protein) are the likely binding site(s) for MTH0777 (*51, 52*). Since these domains (KTN and RCK) are found mainly in uptake or efflux systems (transport systems) we propose that MTH0776 and MTH0777 form complexes that lead to the transport of particular molecules/ligands which is specific to the methanogenic archaebacteria.

In summary we present the solution structure of MTH0776, a functionally unknown protein from *Methanobacterium thermoautotrophicum* ΔH, which exhibits a novel fold. Preliminary NMR studies and gene analysis show that MTH0776 and MTH0777 form a two-protein operon system and bind to each other in solution. However, more work needs to be done to completely characterize the interaction between MTH0776 and MTH0777.

4.5 References

1. Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D. and Reeve, J. N. (1997) *J. Bacteriol.*, **179**, 7135 – 7155.
2. Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Le, B., Ramelot, T., Lee, G.M., Bhattacharyya, S., Gutierrez, P., Denisov, A., Lee, C.H., Cort, J.R., Kozlov, G., Liao, J., Finak, G., Chen, L., Wishart, D., Lee, W., McIntosh, L.P., Gehring, K., Kennedy M.A., Edwards, A.M. and Arrowsmith, C.H. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 1825 – 1830.
3. Holzinger, A., Philips, K.S. and Weaver, T.E. (1996) *BioTechniques*, **20**, 804 – 808.
4. Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J.R., Booth, V., Mackereth, C.D., Saridakis, V., Ekiel, I.., Kozlov, G., Maxwell, K.L., Wu. N., McIntosh, P., Gehring, K., Kennedy, M.A., Davidson, A.R., Pai, F.E., Gerstein, M., Edwards, A.M. and Arrowsmith, C.H. (2000) *Nat. Struct. Biol.*, **7**, 903 – 909.
5. Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277 – 293.
6. Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995) *J. Biomol. NMR*, **6**, 135 – 140.
7. Kay, L.E., Xu, Y.G. and Yamazaki, T. (1994) *J. Magn. Reson.*, **109**, 129 – 133.
8. Grzesiek, S. and Bax, A. (1992) *J. Am. Chem. Soc.*, **114**, 6291.
9. Zhang, O., Kay, L.E., Olivier, J.P. and Forman-Kay, J.D. (1994) *J. Biol. NMR*, **4**,

845 - 858.

10. Kuboniwa, H., Grzesiek, S., Delagio, F. and Bax, A. (1994) *J. Biomol. NMR*, **4**, 871 – 878.

11. Grzesiek, S., Anglister, J. and Bax, A. (1993) *J. Magn. Reson.*, **101**, 114 – 119.

12. Amegbey, G., Chang, Z., Stothard, P., Yee, A., Arrowsmith, C. and Wishart, D.S. (2004) *J.Biomol. NMR*, **30**, 459 – 460.

13. Wuthrich, K. (1986) NMR of Proteins and Nucleic Acids, John Wiley and Sons Inc. New York

14. Cornilescu, F., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR*, **13**, 289 – 302.

15. Wishart, D. S. and Sykes, B.D. (1994) *Meth. Enzymol.*, **239**, 363 – 392.

16. Schwieters, C.D., Kuszewski, J.J., Tjiandra, N. and Clore, G.M. (2003) *J. Magn. Res.*, **160**, 66 – 74.

17. Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R., and Thornton, J. (1996) *J. Biomol. NMR*, **8,** 477 – 486.

18. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graphics*, **14**, 51-55

19. Maiti, R., Van Domselaar, G., Zhang, H.and Wishart, D. (2004) *Nucl. Acids Res.*, **32**, W590 – W594.

20. Altschul, S.E., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) *J. Mol. Biol.*, **215**, 403 – 410.

21. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmorgin, F. and Higgins, D.G. (1997) *Nuc. Acids Res.*, **25**, 4876 – 4882.

22. Stothard, P. and Wishart, D.S. (2005) *Bioinformatics*, **21**, 537 – 539.

23. Saitou, N. and Nei, M. (1987) *Mol. Biol. Evol.*, **4**, 406 – 425.

24. Sayle, R.A. and Milner-White, E.J. (1995) *Trends Biochem. Sci.*, **20**, 374.

25. Xu, X.P. and Case, D.A. (2001) *J. Biomol. NMR*, **21**, 321 – 333.

26. Willard, L., Anuj, R., Zhang, H., Monzavi, H., Boyko, R., Sykes, B.D. and Wishart, D.S. (2003) *Nucl. Acids Res.*, **31**, 3316 – 3319.

27. Shindyalov, I.N., and Bourne, P.E. (1998) *Protein Eng.*, **11**, 739 – 747.

28. Holm, L. and Sander, C. (1996) *Science*, **273**, 595 – 602.

29. Lo Conte L., Brenner S.E., Hubbard T.J.P., Chothia C., Murzin A. (2002) *Nucl. Acid Res.*, **30**, 264 – 267.

30. Yee, A., Pardee, K., Christendat, D., Savchenko, A., Edwards, A.M. and Arrowsmith, C.H. (2003) *Acc. Chem. Res.*, **36**, 183-189.

31. Govindarajan, S., Recabarren, R. and Goldstein, R.A. (1999) *Proteins*, **35**, 408 – 414.

32. Orengo, C.A., Bray, J.E., Buchan, D.W.A., Harison, A., Lee, D., Pearl, F.M.G., Sillitoe, I., Todd, A.E., and Thornton, J.M. (2002) *Proteomics*, **2**, 11 – 21.

33. Gibrat, J.F., Madej, T. and Bryant, S.H. (1996) *Curr. Opin. Struct. Biol.*, **6**, 377 – 385.

34. Hou, J., Jun, S-R., Zhang, C. and Kim, S-H. (2005) *PNAS*, **102**, 3651 – 3656.

35. Willard, L., Anuj, R., Zhang, H., Monzavi, H., Boyko, R., Sykes, B.D. and Wishart, D.S. (2003) *Nucl. Acids Res.*, **31**, 3316 – 3319.

36. Hooft, R.W.W., Vriend, G., Sander, C. and Abola, E.E. (1996) *Nature*, **381**, 272.

37. Roosild, T.P., Miller, S., Booth, I.R. and Choe, S. (2002) *Cell*, **109**, 781 – 791.

38. Bellamacina, C.R. (1996) *FASEB J.*, **10**, 1257 – 1269.

39. Jiang, Y., Pico, A., Cadene, M., Chait, B.T. and Mackinnon, R. (2001) *Neuron*, **29**,

593 – 601.

40. Branden, C. and Tooze, J. (1991) In Introduction to Protein Stucture: Enzymes that bind Nucleotides, Branden, C. and Tooze, J. (Eds) New York Garland Publishing, pp141 – 159.

41. Parra-Lopez, C., Lin, R., Aspedon, A. and Groisman, E.A. (1994) *EMBO J.*, **13**, 3964 – 3972.

42. Ferguson, G.P., Munro, A.W., Douglas, R.M., McLaggan, D. and Booth, I.R. (1993) *Mol. Microbiol.*, **9**, 1297 – 1303.

43. Schlosser, A., Hamann, A., Bossemeyer, D., Schneider, E. and Bakker, E.P. (1993) *Mol. Microbiol.*, **9**, 533 – 543.

44. Rhoads, D.B and Epstein, W. (1977) *J. Biol. Chem.*, **252**, 1394 – 1401.

45. Stewart, L.M.D., Bakker, E.P. and Booth, I.R. (1985) *J. Gen. Microbiol.*, **131**, 77 – 85.

46. Alberts, B. (1998) *Cell*, **92**, 291 – 294.

47. Oliver, S. (2000) *Nature*, **403**, 601 – 603.

48. Uetz, P., Glot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emill, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S. and Rothberg, J.M. (2000) *Nature*, **403**, 623 – 627.

49. Laskowski, R.A., Watson, J.D. and Thornton, J.M (2005) *Nuc. Acids Res.*, **33**, W89 – W93.

50. Binkowski, T.A., Naghibzadeh, S. and Liang, J. (1998) *Nuc. Acids Res.*, **31**, 3352 – 3355.

51. Watson, J.D., Laskowski, R.A. and Thornton, J.M (2005) *Curr. Opin. Struct. Biol.*, **15**, 275 – 284.

52. Campbell, S.J., Gold, N.D., Jackson, R.M. and Westhead, D.R. (2003) *Curr. Opin. Struct. Biol.*, **13**, 389 – 395.

Table 4.1 Structural statistics for the ensemble of 20 structures for MTH0776

Distance Restraints	
All	1844
Intra residue	146
Sequential ($\vert i - j \vert = 1$	698
Medium range ($2 \leq \vert i - j \vert \leq 5$)	422
Long range ($5 < \vert i - j \vert$)	578
Dihedral angle restraints	
All	164
Phi (Φ)	84
Psi (Ψ)	80
Hydrogen bonds	40 X 2
RMSD from ideal geometry	
Bonds (Å)	0.00258 ± 0.00005
Angles (deg)	0.5865 ± 0.0047
Impropers (deg)	0.4525 ± 0.0033
RMSD from experimental data	
NOES (Å)	0.0461 ± 0.0012
Dihedrals (deg)	0.7216 ± 0.0238
Ramachandran plot analysis	
Most favoured	72.9%
Allowed	25.2%
Generously allowed	1.9%
Disallowed	0.1%
RMSD from the mean structure (Å)	
All residues	
Backbone	0.51 ± 0.12
All heavy atoms	1.49 ± 0.13
Equivalent Resolution for Stereochemical Parameters (Å)	
Percentage residues in A, B, L	2.7
H-bond energy	1.8
Chi-1 pooled	2.4
Chi-2 trans	2.5

Figure 4.01 pET-15b cloning/expression system from Novagen. The MTH0776 gene was sub-cloned into the multiple cloning sites such that the protein is expressed with an additional 20 amino acids at the N-terminus

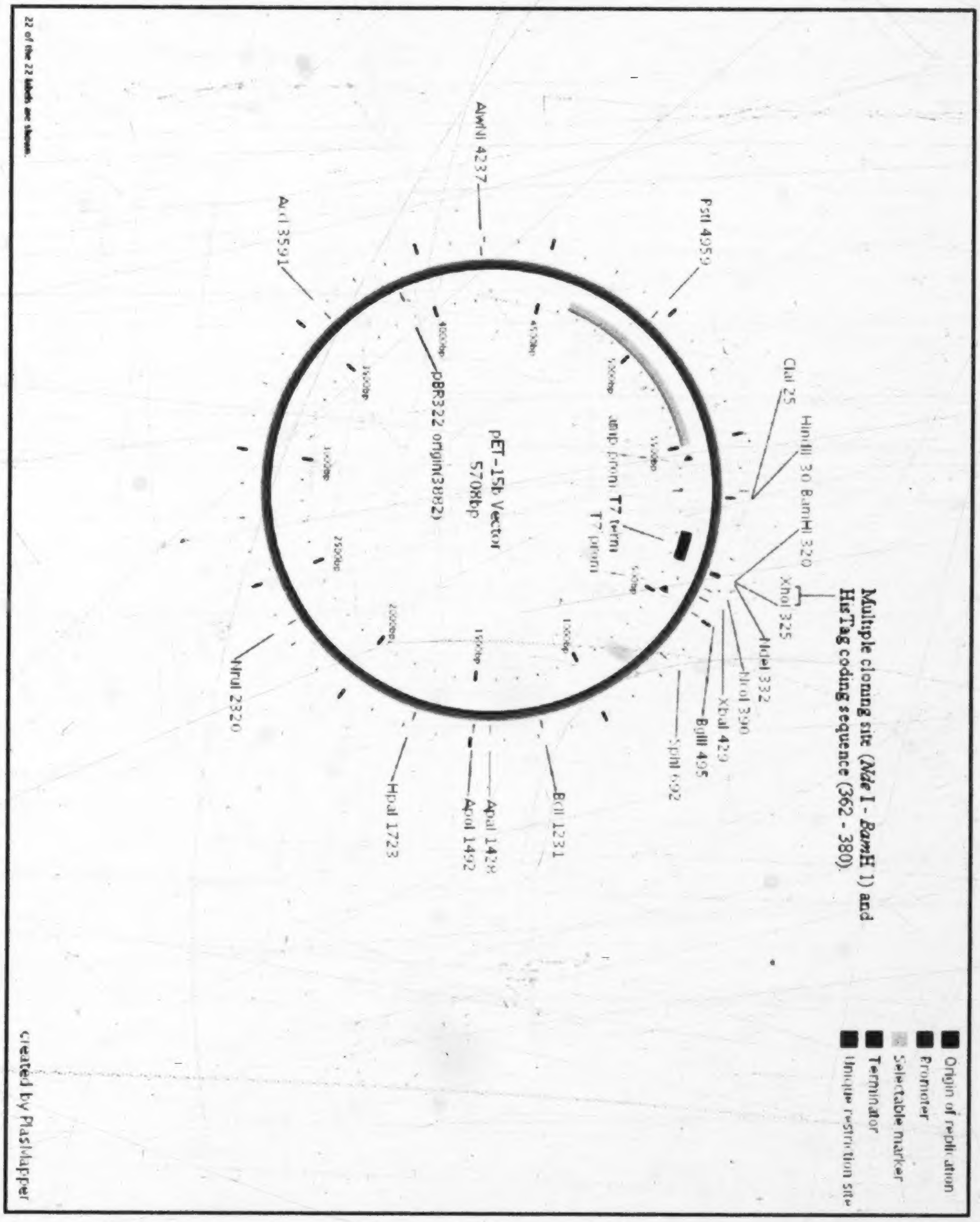

Multiple cloning site (*Nde* I - *BamH* 1) and HisTag coding sequence (362 - 380).
ClaI 25
HindIII 30
BamHI 320
XhoI 325
NdeI 332
NcoI 390
XbaI 429
BglII 495
SphI 692
BclI 1231
ApaI 1428
ApoI 1492
HpaI 1723
NruI 2320
AvrII 3591
AlwNI 4237
PstI 4959
pBR322 origin(3882)
amp prom
T7 term
T7 prom
pET-15b Vector
5708bp
Origin of replication
Promoter
Selectable marker
Terminator
Unique restriction site
22 of the 22 labels are shown.
created by PlasMapper

Figure 4.02 ^{1}H-^{15}N HSQC spectrum from 1 mM MTH 0776 and 2 mM MTH0777 in 700 μL of a buffer solution pH 6.8, made up of 50 mM NaH_2PO_4, 300 mM NaCl, 50 μL D_2O, 1 mM DSS, 15 mM DTT and 10 μL of a 3% solution of sodium azide, collected on a Varian Inova 500 MHz spectrometer. (Spectrum is rendered at a contour level ¼ that of figure 4.03.

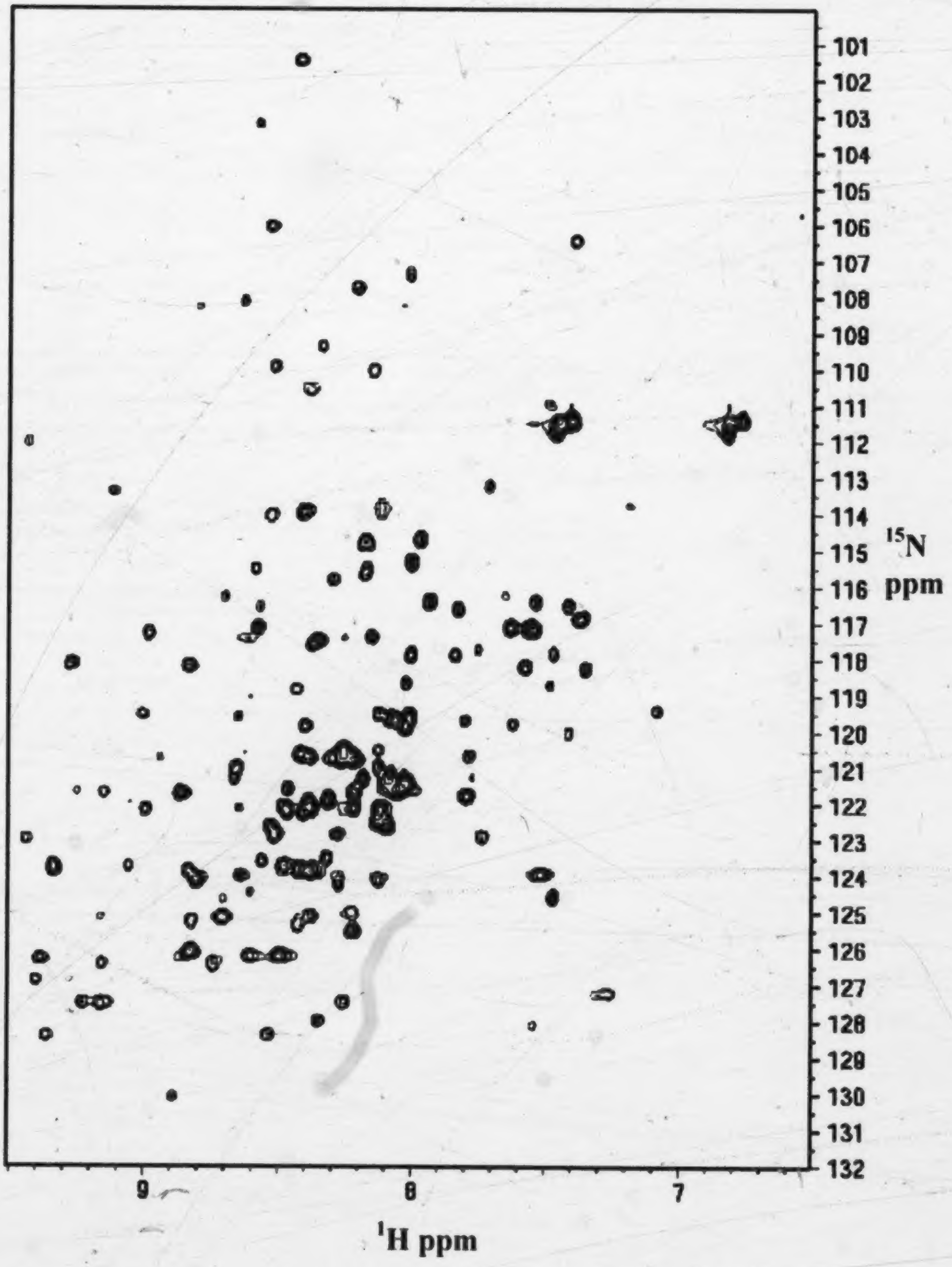


Figure 4.03 ^{1}H-^{15}N HSQC spectrum from 1 mM MTH 0776 in 500 μL of a buffer solution pH 6.8, made up of 50 mM NaH_2PO_4, 300 mM NaCl, 50 μL D_2O, 1 mM DSS, 15 mM DTT and 10 μL of a 3% solution of sodium azide, collected on a Varian Inova 500 MHz spectrometer.

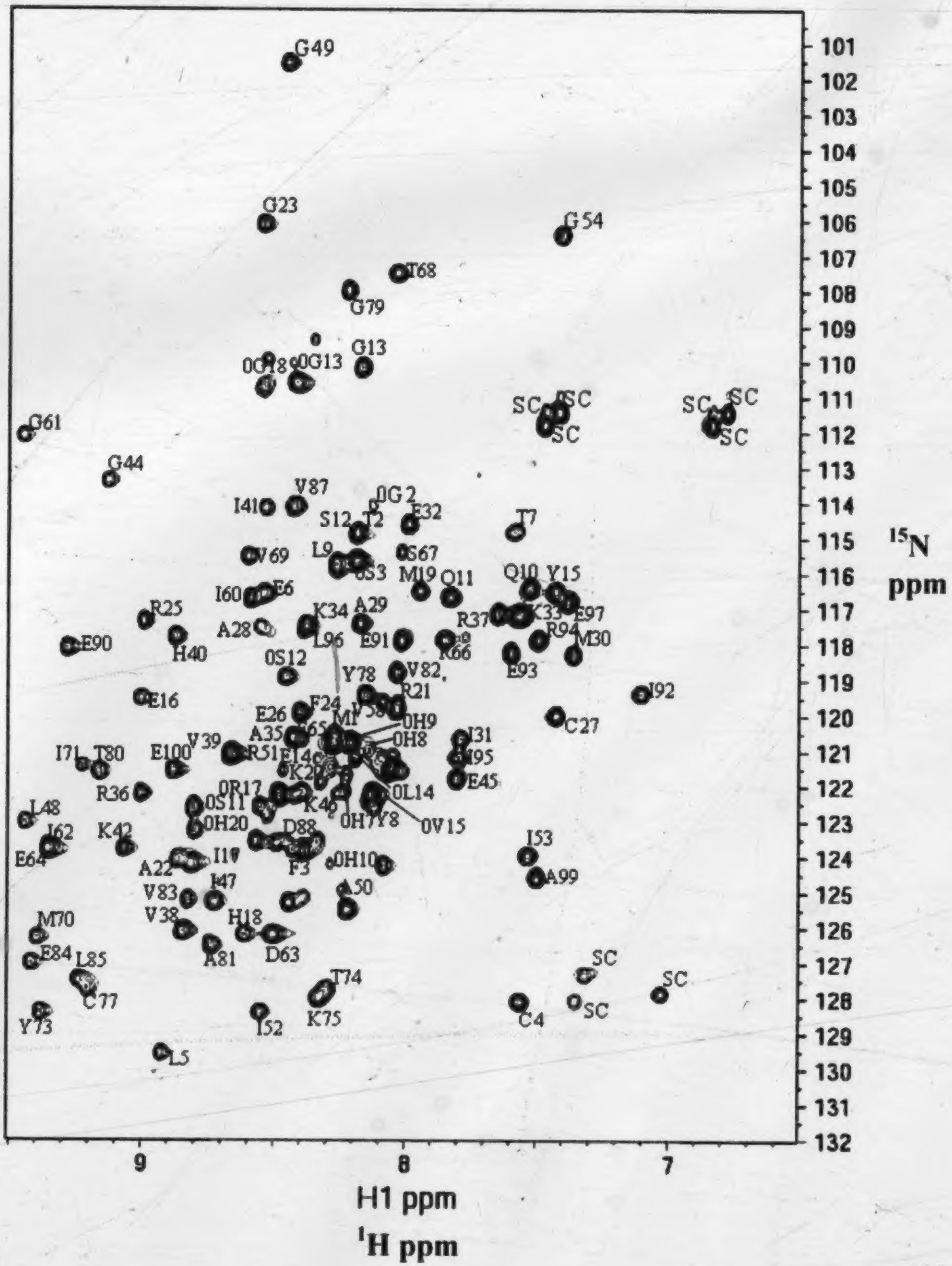
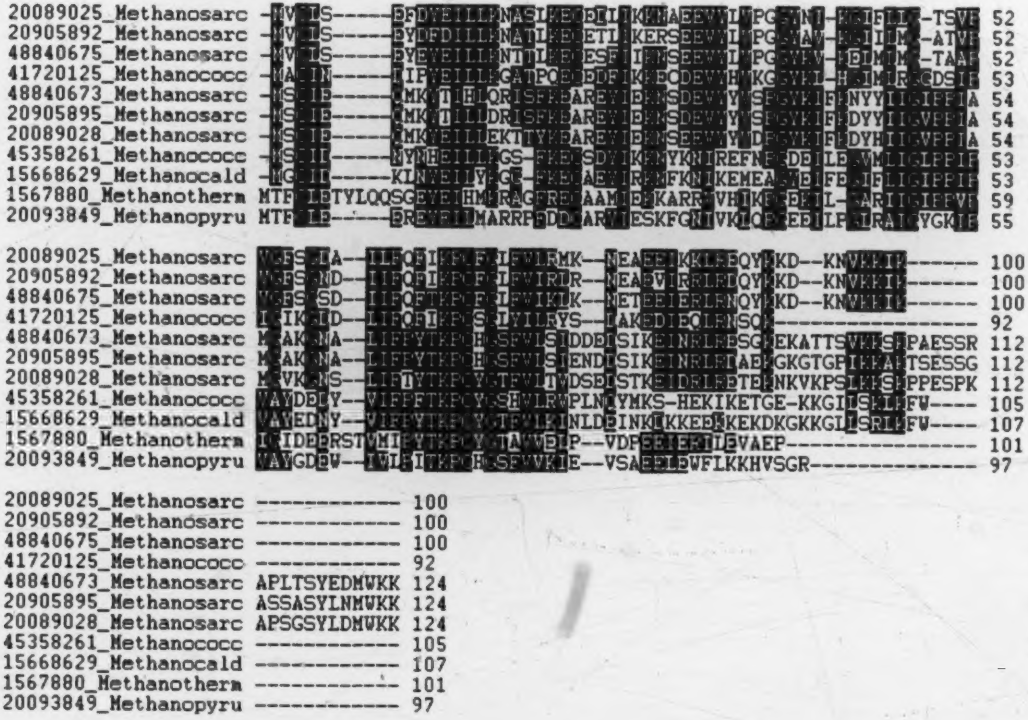


Figure 4.04 A BLAST search for proteins related to MTH0776

gi|15678801|ref|NP_275918.1|	conserved protein [Methanother...	160	2e-40
gi|20093849|ref|NP_613696.1|	Uncharacterized protein conser.....	60	3e-10
gi|45358261|ref|NP_987818.1|	conserved hypothetical protein.....	47	2e-06
gi|15668629|ref|NP_247427.1|	conserved hypothetical protein.....	45	1e-05
gi|48840673|ref|ZP_00297599.1|	COG4033: Uncharacterized pro...	45	1e-05
gi|48840675|ref|ZP_00297601.1|	COG4033: Uncharacterized pro...	43	5e-05
gi|20905895|gb|AAM31111.1|	conserved protein [Methanosarcin...	41	1e-04
gi|20089025|ref|NP_615100.1|	conserved hypothetical protein......	41	2e-04
gi|20089028|ref|NP_615103.1|	conserved hypothetical protein......	39	5e-04
gi|41720125|ref|ZP_00148964.1|	COG4033: Uncharacterized pro...	35	0.013
gi|20905892|gb|AAM31109.1|	conserved protein [Methanosarcin...	34	0.025

Figure 4.05 Alignment made using ClustalX using the default settings. Residues identical among 50% or more of the sequences are colored black. Residues similar among 50% or more of the sequences are colored gray.



```
20089025_Methanosarc ------------- 100
20905892_Methanosarc ------------- 100
48840675_Methanosarc ------------- 100
41720125_Methanococc ------------- 92
48840673_Methanosarc APLTSYEDMWKK 124
20905895_Methanosarc ASSASYLNMWKK 124
20089028_Methanosarc APSGSYLDMWKK 124
45358261_Methanococc ------------- 105
15668629_Methanocald ------------- 107
1567880_Methanotherm ------------- 101
20093849_Methanopyru ------------- 97
```

Figure 4.06

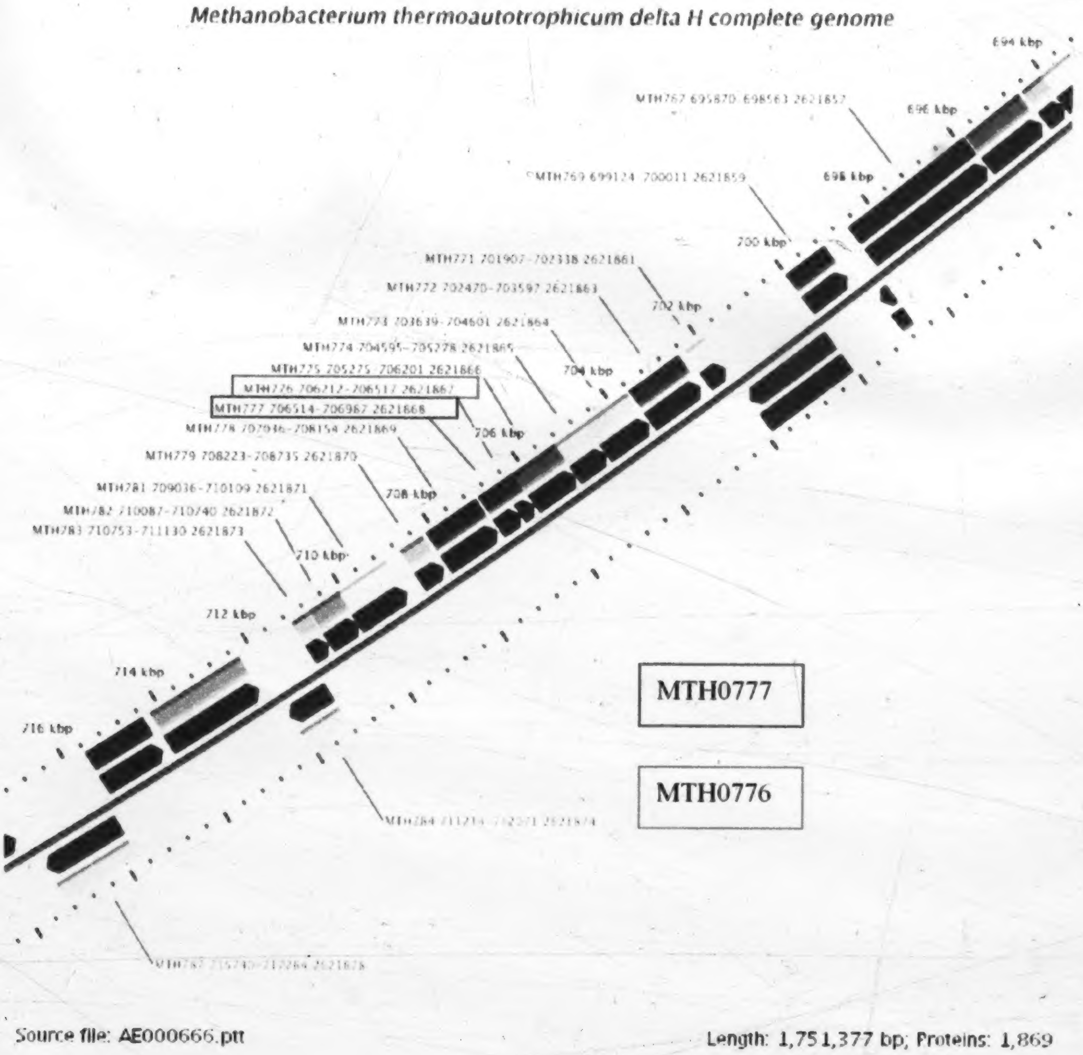
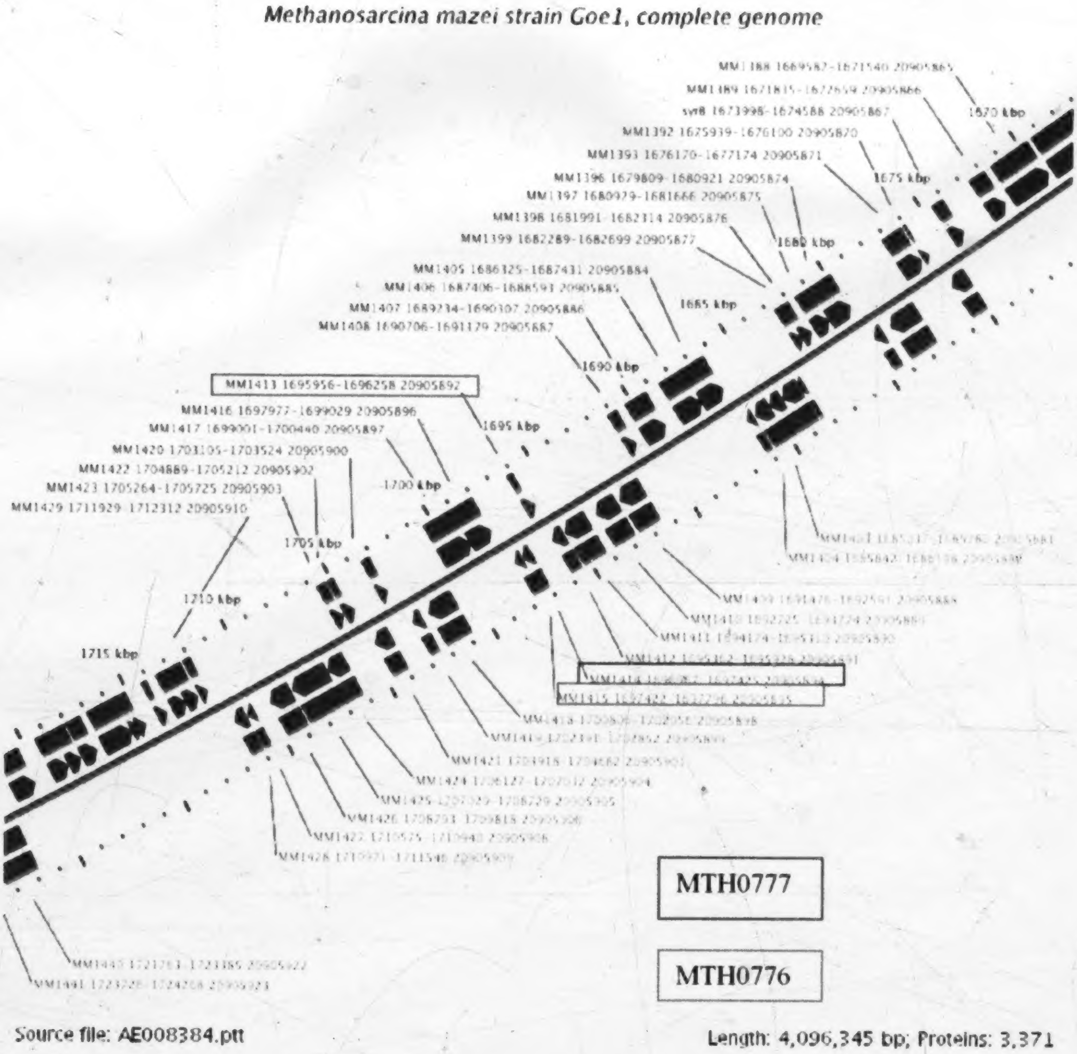
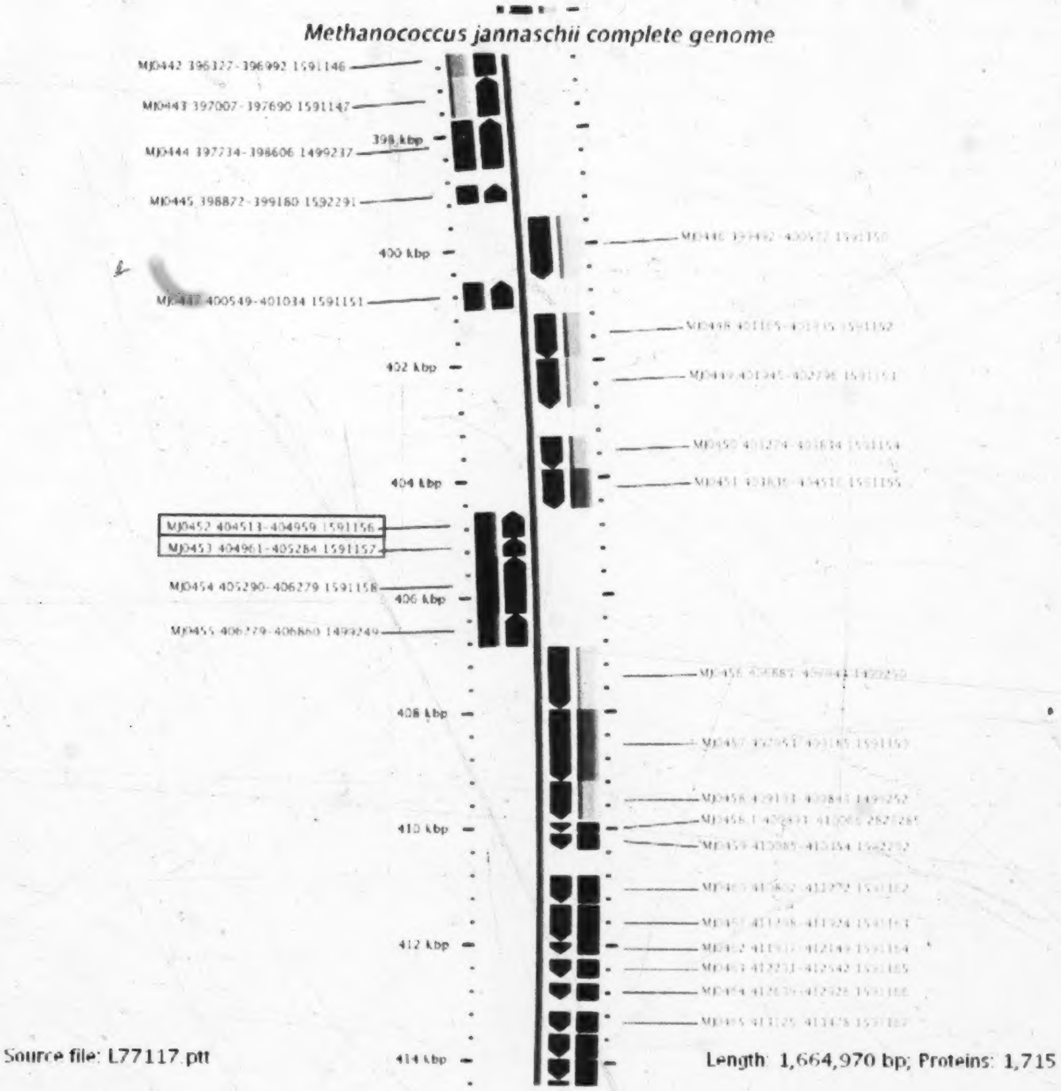

Methanobacterium thermoautotrophicum delta H complete genome
694 kbp
MTH767 695870-698563 2621857
696 kbp
MTH769 699124-700011 2621859
698 kbp
700 kbp
MTH771 701907-702338 2621861
MTH772 702470-703597 2621863
702 kbp
MTH773 703639-704601 2621864
MTH774 704595-705278 2621865
704 kbp
MTH775 705275-706201 2621866
MTH776 706212-706517 2621867
MTH777 706514-706987 2621868
706 kbp
MTH778 707046-708154 2621869
MTH779 708223-708735 2621870
708 kbp
MTH781 709036-710109 2621871
MTH782 710087-710740 2621872
MTH783 710753-711130 2621873
710 kbp
712 kbp
714 kbp
716 kbp
MTH0777
MTH0776
MTH784 711214-712271 2621874
MTH787 715740-717264 2621876
Source file: AE000666.ptt
Length: 1,751,377 bp; Proteins: 1,869

Figure 4.07


Methanosarcina mazei strain Goe1, complete genome
MM1388 1669582-1671540 20905865
MM1389 1671815-1672659 20905866
svtB 1673998-1674588 20905867
MM1392 1675939-1676100 20905870
MM1393 1676170-1677174 20905871
MM1396 1679809-1680921 20905874
MM1397 1680929-1681666 20905875
MM1398 1681991-1682314 20905876
MM1399 1682289-1682699 20905877
MM1405 1686325-1687431 20905884
MM1406 1687406-1688591 20905885
MM1407 1689234-1690307 20905886
MM1408 1690706-1691129 20905887
MM1413 1695956-1696258 20905892
MM1416 1697977-1699029 20905896
MM1417 1699001-1700440 20905897
MM1420 1703105-1703524 20905900
MM1422 1704889-1705212 20905902
MM1423 1705264-1705725 20905903
MM1429 1711929-1712312 20905910
1670 kbp
1675 kbp
1680 kbp
1685 kbp
1690 kbp
1695 kbp
1700 kbp
1705 kbp
1710 kbp
1715 kbp
MM1401 1685317-1685780 20905881
MM1404 1685842-1686136 20905883
MM1409 1691476-1692591 20905888
MM1410 1692725-1693724 20905889
MM1411 1694174-1695312 20905890
MM1412 1695362-1695926 20905891
MM1414 1696087-1697425 20905894
MM1415 1697422-1697796 20905895
MM1418 1700606-1702051 20905898
MM1419 1702391-1702852 20905899
MM1421 1703918-1704662 20905901
MM1424 1706127-1707512 20905904
MM1425 1707529-1708729 20905905
MM1426 1708751-1709818 20905906
MM1427 1710525-1710940 20905908
MM1428 1710971-1711546 20905909
MM1440 1721763-1723185 20905922
MM1441 1723728-1724208 20905923
MTH0777
MTH0776
Source file: AE008384.ptt
Length: 4,096,345 bp; Proteins: 3,371

Note that the organism has a duplicated gene of MTH0776

Figure 4.08

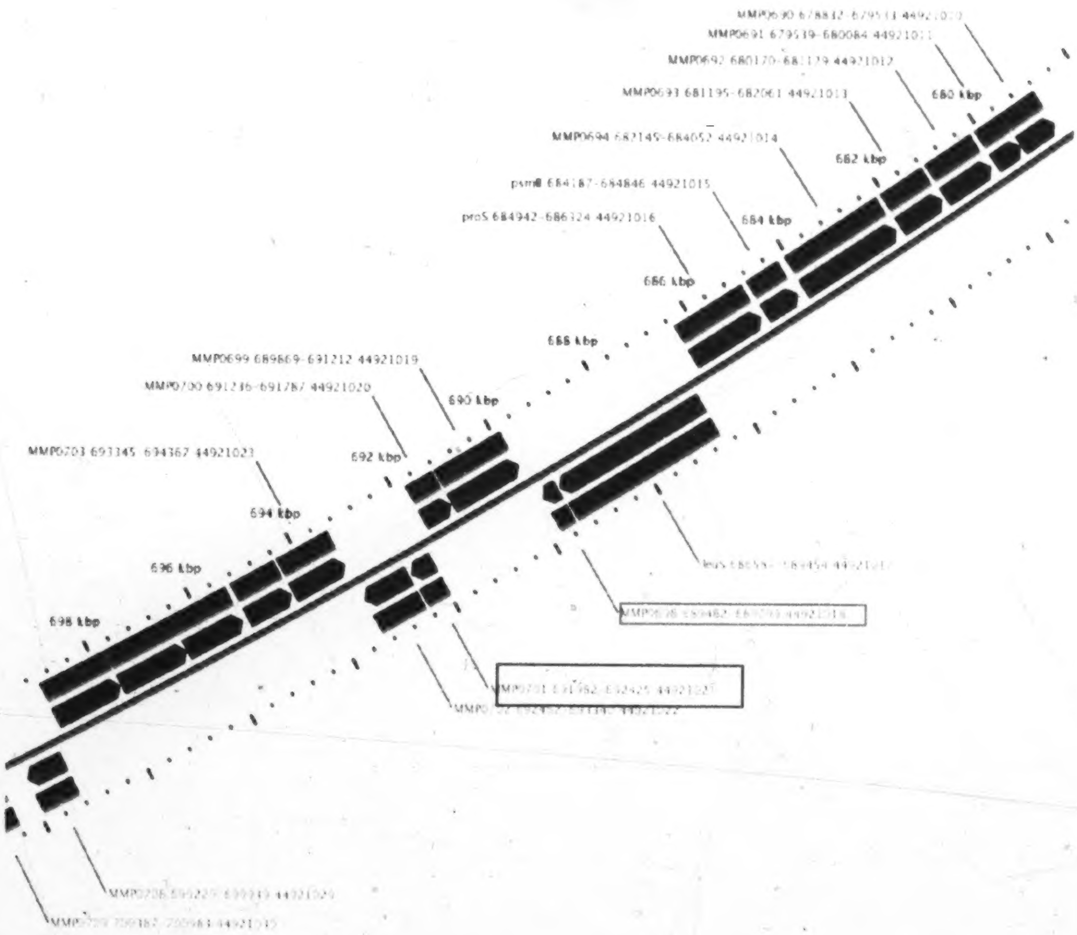

Methanococcus jannaschii complete genome
MJ0442 396127-396992 1591146
MJ0443 397007-397690 1591147
398 kbp
MJ0444 397734-398606 1499237
MJ0445 398872-399180 1592291
400 kbp
MJ0446 399492-400572 1591150
MJ0447 400549-401034 1591151
MJ0448 401105-401315 1591152
402 kbp
MJ0449 401345-402296 1591153
MJ0450 403274-403834 1591154
404 kbp
MJ0451 403835-404510 1591155
MJ0452 404513-404959 1591156
MJ0453 404961-405284 1591157
MJ0454 405290-406279 1591158
406 kbp
MJ0455 406279-406860 1499249
MJ0456
408 kbp
MJ0457
MJ0458
410 kbp
MJ0458.1
MJ0459
412 kbp
MJ0461
MJ0462
MJ0463
MJ0464
MJ0465
414 kbp
Source file: L77117.ptt
Length: 1,664,970 bp; Proteins: 1,715

MTH0777

MTH0776

Figure 4.09

Methanococcus maripaludis strain S2, complete sequence.


MMP0690 678832-679513 44921010
MMP0691 679539-680084 44921011
MMP0692 680170-681129 44921012
MMP0693 681195-682061 44921013
MMP0694 682145-684052 44921014
psmB 684187-684846 44921015
proS 684942-686324 44921016
680 kbp
682 kbp
684 kbp
686 kbp
688 kbp
690 kbp
692 kbp
694 kbp
696 kbp
698 kbp
MMP0699 689869-691212 44921019
MMP0700 691236-691787 44921020
MMP0703 693345-694367 44921023
MMP0701 691982-692425 44921021
leuS
MMP0698
MMP0702
MMP0706
MMP0709

Source file: BX950229.ptt

Length: 1,661,137 bp, Proteins: 1,722

MTH0777

MTH0776

Figure 4.10

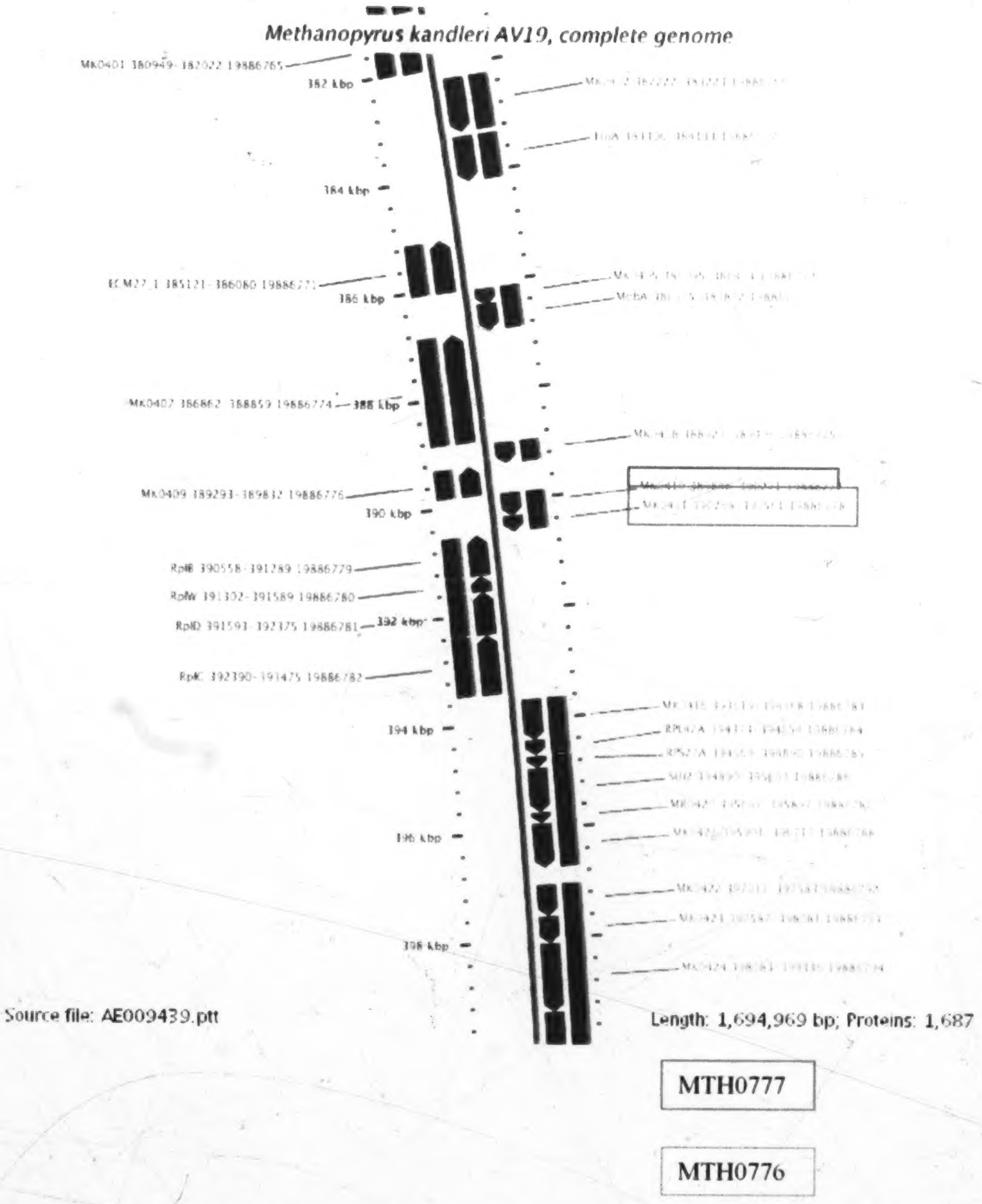
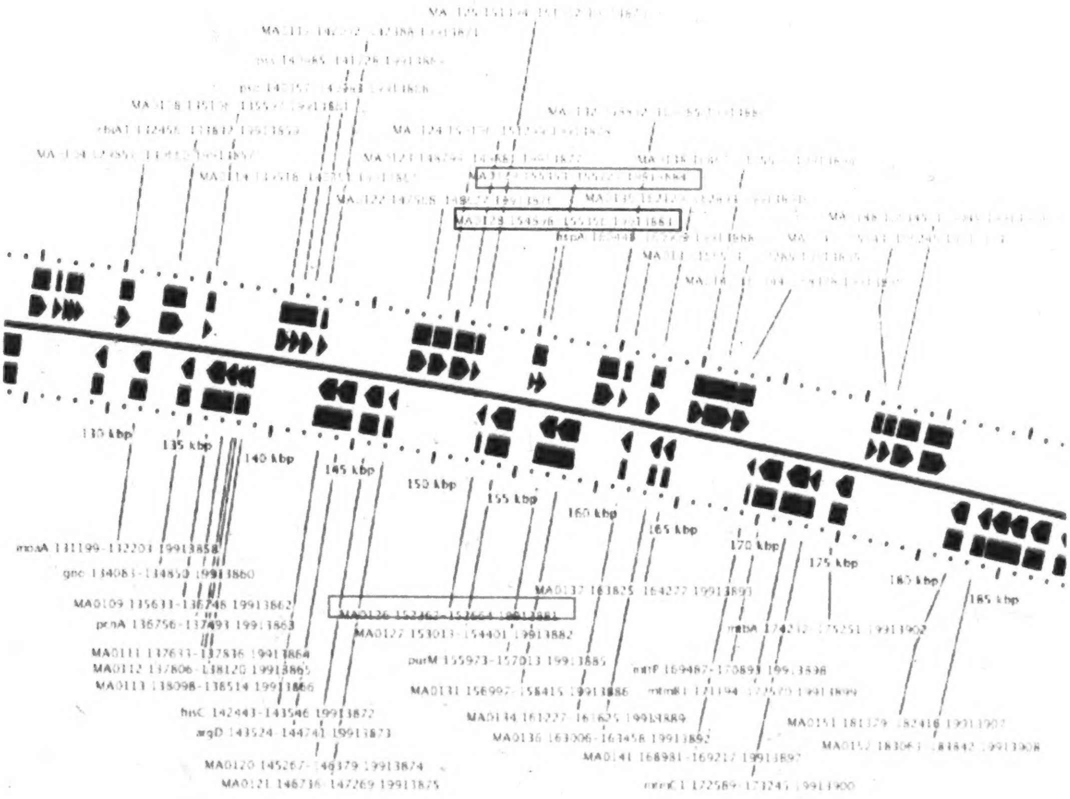

Methanopyrus kandleri AV19, complete genome
MK0401 380949-382022 19886765
382 kbp
384 kbp
ECM27_1 385121-386080 19886771
386 kbp
MK0407 386862-388859 19886774
388 kbp
MK0409 389293-389832 19886776
390 kbp
RplB 390558-391289 19886779
RplW 391302-391589 19886780
RplD 391593-392375 19886781
392 kbp
RplC 392390-393475 19886782
394 kbp
396 kbp
398 kbp
Source file: AE009439.ptt
Length: 1,694,969 bp; Proteins: 1,687
MTH0777
MTH0776

Figure 4.11

Methanosarcina acetivorans str. C2A, complete genome


130 kbp
135 kbp
140 kbp
145 kbp
150 kbp
155 kbp
160 kbp
165 kbp
170 kbp
175 kbp
180 kbp
185 kbp
moaA 131199-132203 19913858
gnc 134083-134850 19913860
MA0109 135631-136748 19913862
pcnA 136756-137493 19913863
MA0111 137633-137836 19913864
MA0112 137806-138120 19913865
MA0113 138098-138514 19913866
hisC 142443-143546 19913872
argD 143524-144741 19913873
MA0120 145267-146379 19913874
MA0121 146736-147269 19913875
MA0126 152362-152664 19913881
MA0127 153013-154401 19913882
purM 155973-157013 19913885
MA0131 156997-158415 19913886
MA0134 161227-161625 19913889
MA0136 163006-163458 19913892
MA0137 163825-164277 19913893
MA0141 168981-169217 19913897
mttP 169487-170893 19913898
mtmB1 171194-172570 19913899
mtmC1 172589-173245 19913900
mtbA 174212-175251 19913902
MA0151 181129-182416 19913907
MA0152 183063-183842 19913908

Source file: AE010299.ptt

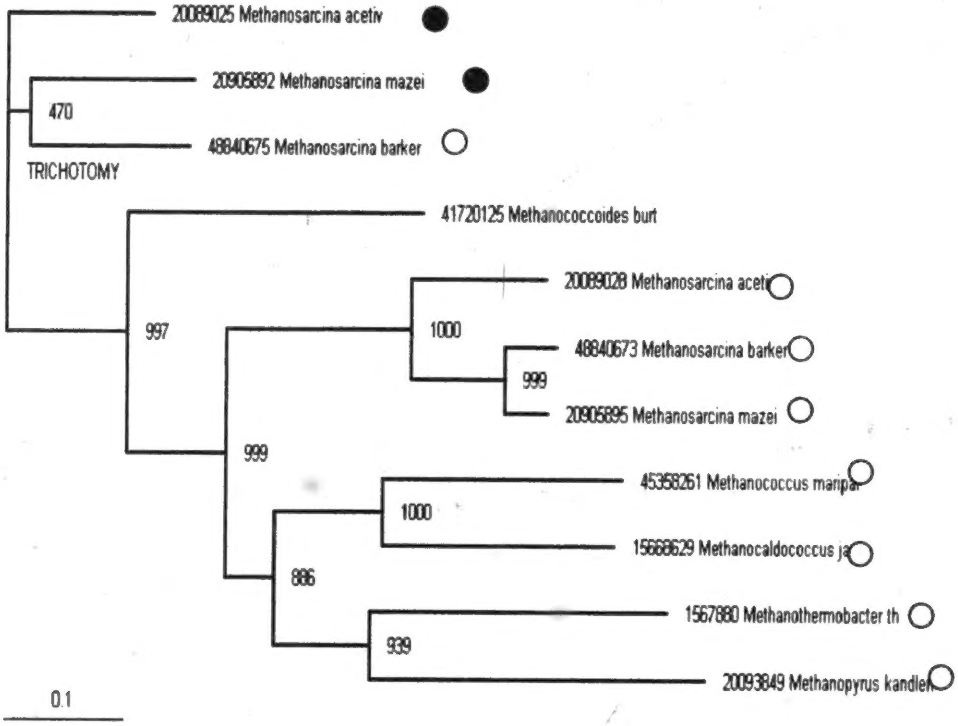
Length: 5,751,492 bp, Proteins: 4,540

MTH0777

MTH0776

Note that the organism has a duplicated gene of MTH0776

Figure 4.12 Phylogenetic analysis of MTH0776 sequences. A phylogenetic tree was generated from aligned sequences using CLUSTALX and the neighbour-joining method. Portions of the alignment containing gaps were excluded from the analysis. The results were analyzed using the bootstrap method (1000 replicates) to provide confidence levels for tree topology. Scale bar: 0.1 changes per site.


20089025 Methanosarcina acetiv
20905892 Methanosarcina mazei
470
48840675 Methanosarcina barker
TRICHOTOMY
41720125 Methanococcoides burt
20089028 Methanosarcina aceti
997
1000
48840673 Methanosarcina barker
999
20905895 Methanosarcina mazei
999
45358261 Methanococcus marip
1000
15668629 Methanocaldococcus ja
886
1567880 Methanothermobacter th
939
20093849 Methanopyrus kandler
0.1

○ Genes next to MTH0777 in sequenced genomes

● Duplicated MTH0776 gene next to MTH0777 in sequenced genomes

○ No genomic data

- BLAST and CGView analysis shown that *Methanosarcina acetivorans*, *Methanosarcina mazei* and *Mathanosarcina bakeri* have an extra MTH0776 gene.
- The phylogenetic tree shows that these three species diverged from a common ancestor after the gene duplication.

Figure 4.13 Superimposed ^{1}H-^{15}N HSQC spectra from (a) 1.5 mM MTH0776 (black) and (b) 1 mM MTH0776 and 2 mM MTH0777 (red) in 700 μL of a buffer solution pH 6.8, made up of 50 mM NaH_2PO_4, 300 mM NaCl, 70 μL D_2O, 1 mM DSS, 15 mM DTT and 10 μL of a 3% solution of sodium azide, collected on a Varian Inova 500 MHz spectrometer. (Spectrum has been optimised for clarity)

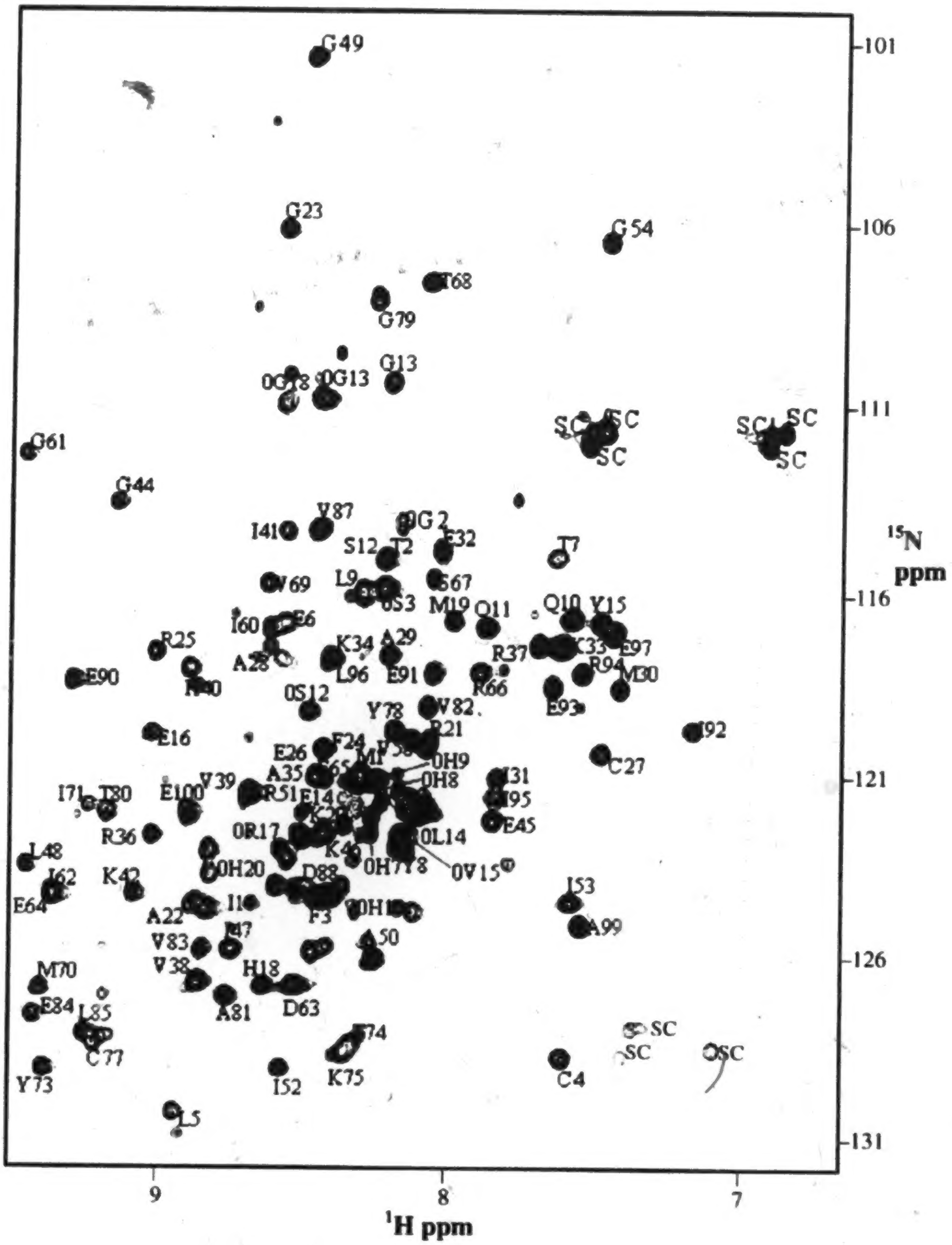


Figure 4.14 CPK model of MTH0776 showing the positions of the residues found to be perturbed in the ^{1}H-^{15}N HSQC spectrum of the mixture of ^{15}N labelled MTH0776 and unlabelled MTH0777.

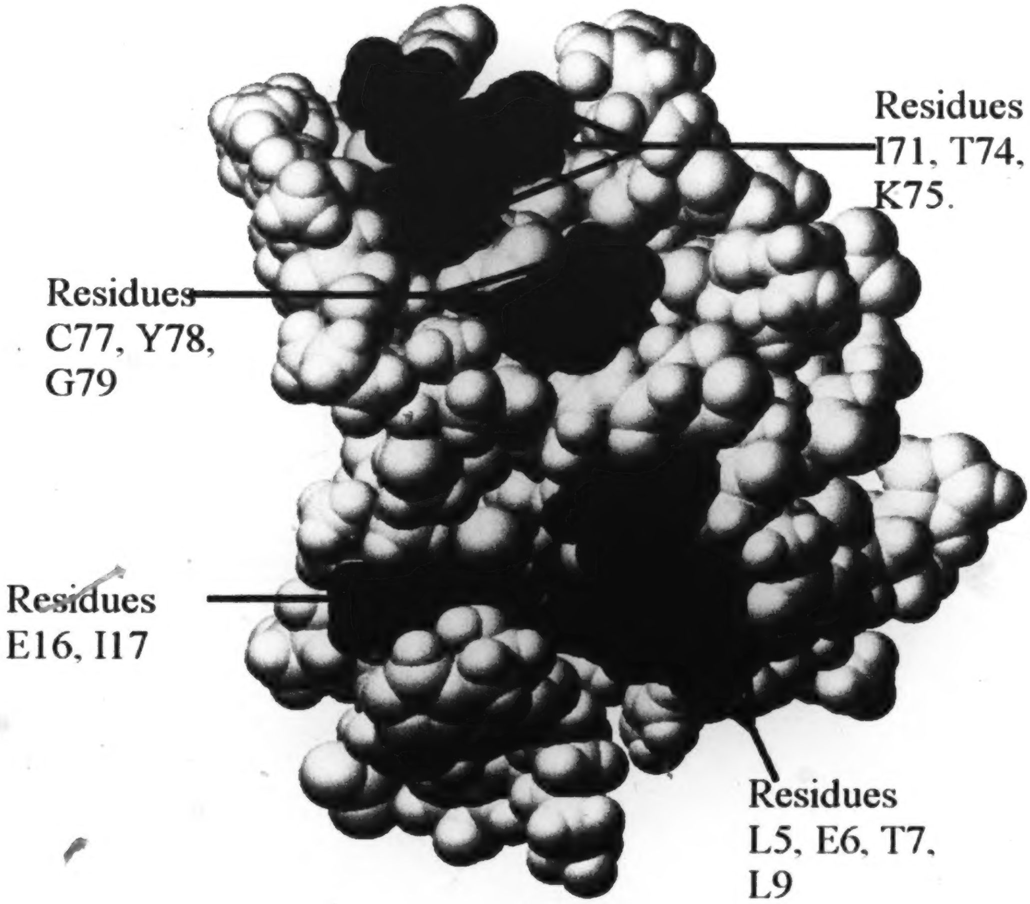

Residues
I71, T74,
K75.
Residues
C77, Y78,
G79
Residues
E16, I17
Residues
L5, E6, T7,
L9

Figure 4.15 Ribbon diagram of MTH0776

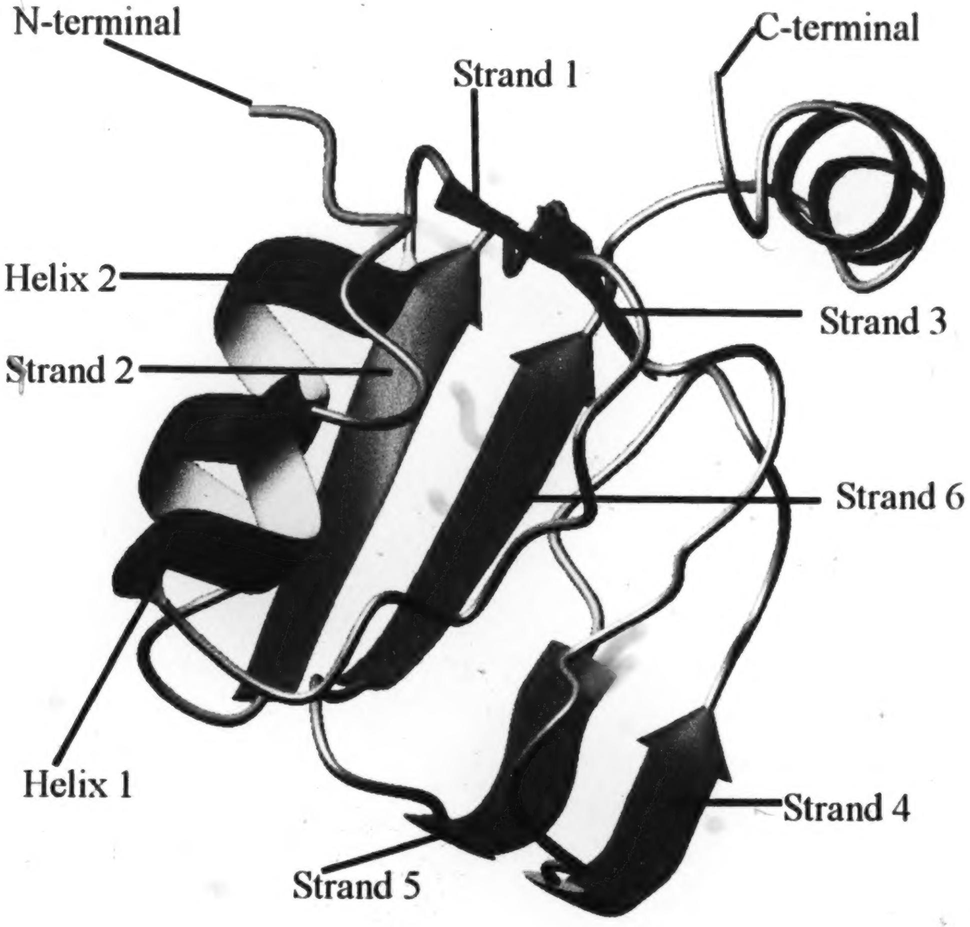

N-terminal
C-terminal
Strand 1
Helix 2
Strand 3
Strand 2
Strand 6
Helix 1
Strand 4
Strand 5

Figure 4.16 Backbone Superposition of the 20 best calculated structures of MTH0776

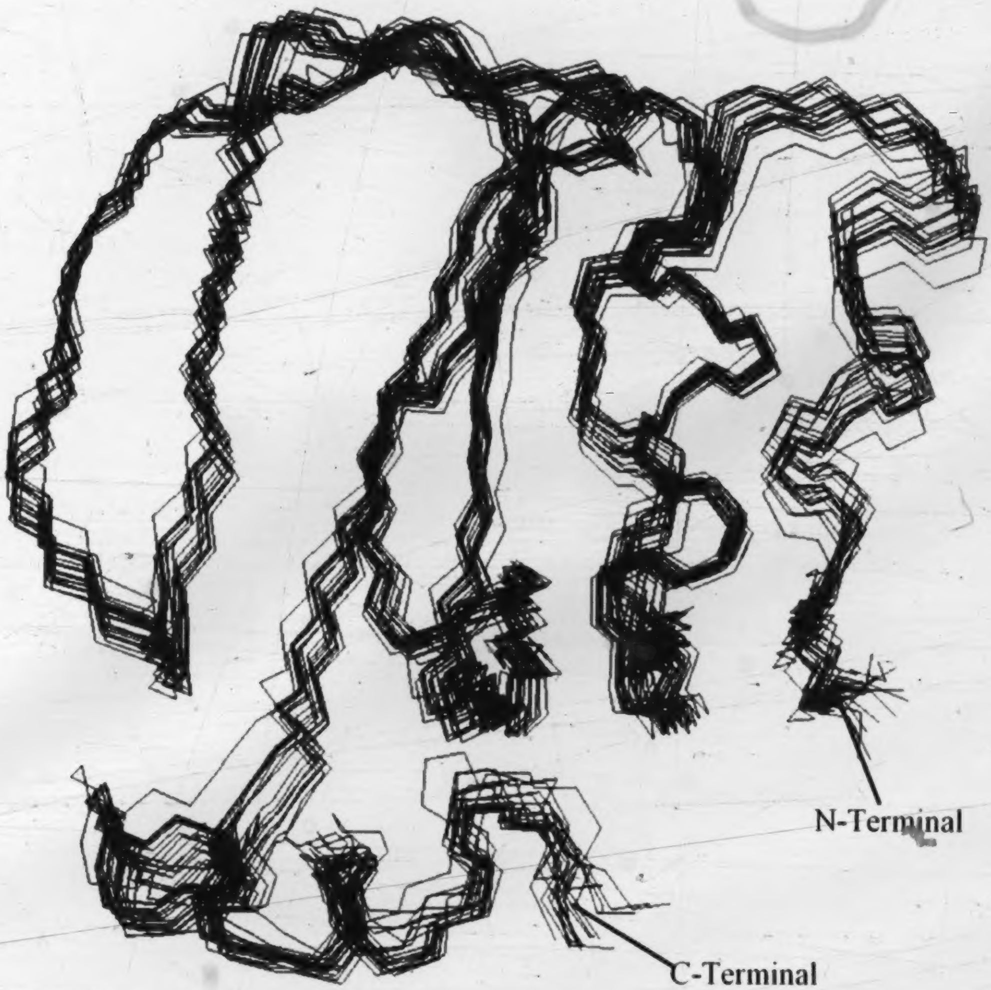
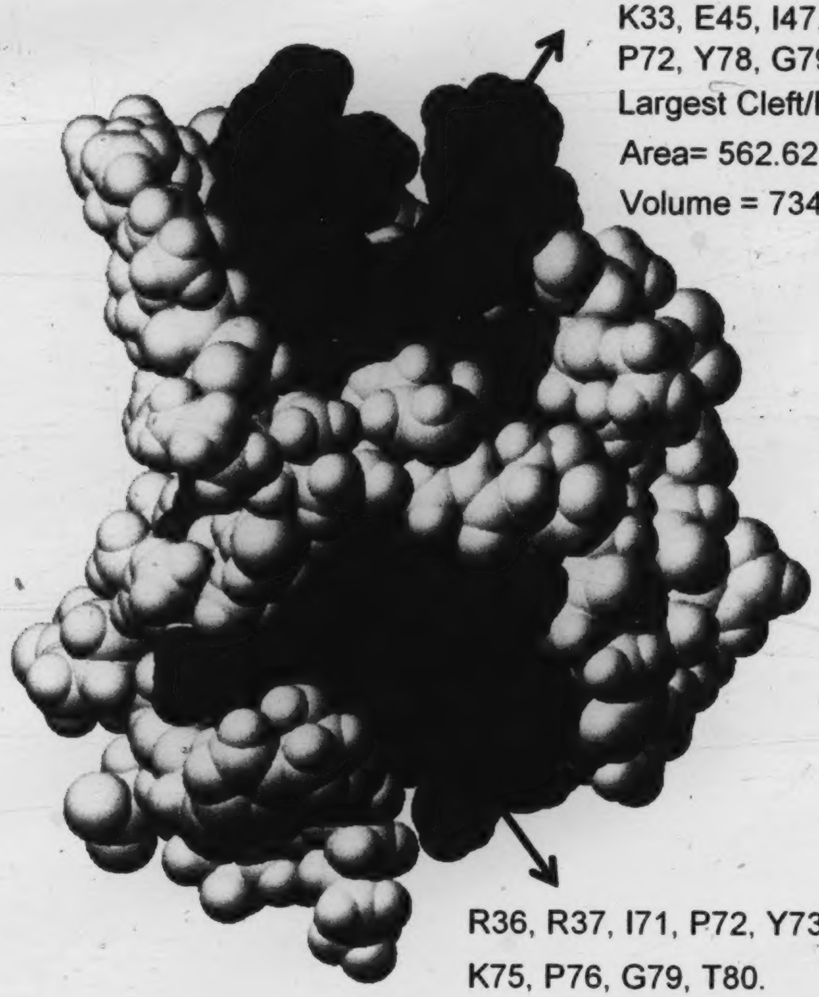

N-Terminal
C-Terminal

Figure 4.17 Residues forming the two largest pocket/cleft on the surface of MTH0776 obtained from ProFunc and Castp webservers



T2, C4, L5, Y8, E14, Y15, E16, R21, A22, G23, E26, M30, K33, E45, I47, R51, M70, P72, Y78, G79, T80, A81, V83

Largest Cleft/Pocket:

Area= 562.62 Å^2

Volume = 734.48 Å^3

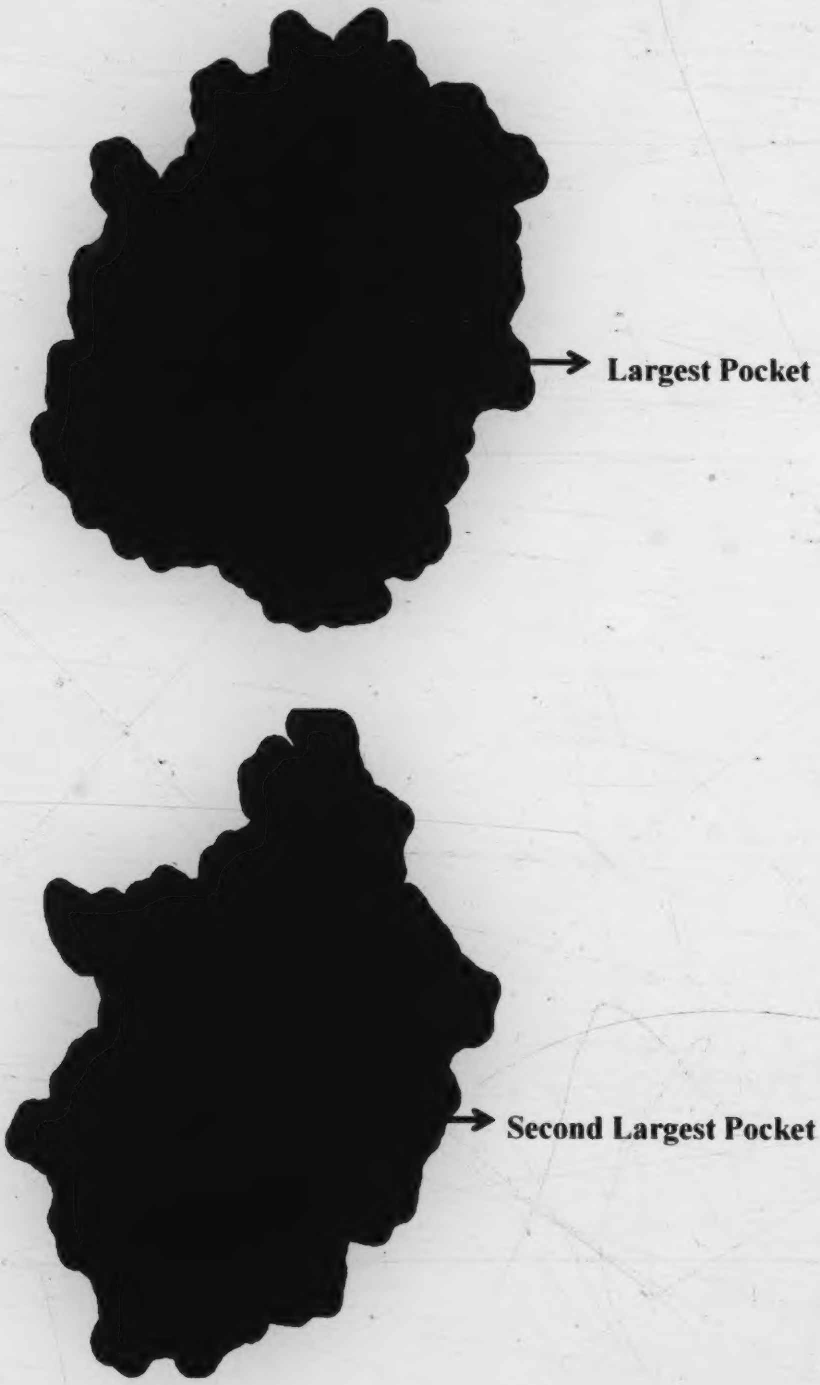
R36, R37, I71, P72, Y73, T74, K75, P76, G79, T80.

Second largest Cleft/Pocket:

Area=144.59 Å^2

Volume= 225.35Å^3

Figure 4.18 Surface rendering with Molmol of the two largest pocket/cleft on the surface of MTH0776 obtained from ProFunc and Castp webservers


Largest Pocket
Second Largest Pocket

Chapter 5

5.1 Conclusions and future studies

The main focus of this thesis has been the determination of protein function through structure. An obvious conclusion from this work which is becoming more common in many structure proteomics work is that when a protein has a novel fold, its function cannot be inferred based on proteins of known structure. However, the fact still remains that knowing the 3D structure of a protein opens up the possibility of ascertaining its function from an analysis of its structure. Recently, many methods have been developed for predicting protein function from structure. These ranges from global comparisons of the proteins structure to other proteins of known 3 D structure to identification of more local features, such as active site residues or DNA-/ligand-binding motifs (*1*). None of these structure-based methods are successful in all cases; therefore a new approach is to use as many methods as possible, both structure-based and sequence-based. This is the principle behind the ProFunc web server (*2, 3*) and I have used this approach extensively in my studies. In chapters 2 and 3, I have shown how the structural details of a protein greatly affect the function it performs. Specifically, we have shown that even though MTH0895 and MTH0807 exhibit a classical glutaredoxin-like fold (and active site sequence in the case of MTH0807), subtle structural details concerning charge distribution, active site geometry and key residue placements on the surface of the proteins lead to an activity more akin to a thioredoxin than a glutaredoxin. This was confirmed by several biochemical tests. In Chapter 2, we also tried to gain some insight into the dual (thioredoxin/glutaredoxin) activity of a related and equally primitive variant of the thioredoxin/glutaredoxin family called T4 glutaredoxin (T4 Grx) using a number of biochemical and biophysical experiments. Our result in combination with work done by previous workers showed that it is theoretically possible to engineer specific properties (reductive and oxidative ability, thioredoxin and/or glutaredoxin activity etc) into redox proteins by changing a few residues at specific positions and around the active site. However, site-directed mutational studies may need to be done on a prototype protein such as MTH0807 or Anabena Trx 2 in order to confirm this hypothesis, particularly with regard to the dual thioredoxin/glutaredoxin activity. This investigation would be particularly useful due to the many potential applications of the NADPH/thioredoxin/thioredoxin reductase systems as drugs and drug targets.

In Chapter 3, I outlined the structural and functional characterization of a novel, ancestral family of thioredoxins found primarily in archeons. Specifically, I described how we used an improved, two-part thioredoxin reductase assay to show that MTH0807 does indeed interact with thioredoxin reductase. In addition, functional and structural comparisons of MTH0807 to an orthologous (51% identity) thioredoxin from the archaebacterium *M. jannaschi* as well as detailed analysis of its own surface topology indicate that MTH0807 is indeed a true thioredoxin and that it belongs to the thioredoxin-thioredoxin reductase system of redox proteins. To further confirm this assessment, we conducted sequence searches against the *M. thermoautotrophicum* genome to attempt to identify possible ferredoxin-thioredoxin homologues and found no significant hits. With this evidence we concluded that MTH0807 is not part of a distinct ribonucleotide-reducing system as had been postulated earlier. However, the unusual activity, sequence and structure of MTH0807 (and its paralogue MTH0895) led us to speculate that these two molecules may represent a group of ancient proteins that were ancestral to both thioredoxins and glutaredoxins.

As was previously described in the first three chapters of this thesis, the NADPH/thioredoxin/thioredoxin reductase system is widely distributed and present in animals as well as plants. It occurs in multiple cell compartments and in plants is housed in the mitochondria, endoplasmic reticulum, cytosol and the seeds. The NADPH/thioredoxin/thioredoxin reductase system also plays numerous cell-critical roles in animals and plants (summarized in Table 1.2). All these actions and regulatory activities indicate that thioredoxins are potential drug targets in combating many ailments that afflict man. For example, it is known that the level of thioredoxin (Trx) in cells is elevated appreciably in many cancers (*4*). When Trx is reduced to its active state by thioredoxin reductase, it enters the nucleus to regulate transcription factor activities. Thioredoxin is also excreted out of the cell where it works with other growth factors to stimulate cell growth. It stimulates the growth of a wide variety of human leukemia and solid tumor lines and when over produced transforms normal cells into cancer cells (*5*). A thioredoxin inhibitor PX-12, has recently completed evaluation in a phase 1 clinical trials (*5*). It is also known that extracellular thioredoxin can associate with plasma membrane and facilitates the recovery of lung function following blockage of the inflow of arterial blood (*6*). Activated thioredoxin has also been found to inactivate neurotoxins and other toxic components in venoms from snakes, bees and scorpions thus suggesting the possibility of using thioredoxins as antidotes against certain venoms (*7*). Recent studies have also implicated thioredoxin as a costimulator in the TNF-induced production of IL-6 and IL-8 in rheumatoid arthritis (*8*).

These results show the striking progress that has been made in our knowledge about thioredoxins. It also shows that thioredoxin fulfills critical functions in both plants and animals, controlling an impressive array of processes fundamental to cell division, cell protection, proliferation and development. These studies have opened the door to further applications of thioredoxin to address a number of problems in technology, nutrition and medicine.

While most of the proposed medical applications for thioredoxins are limited to vertebrate or human forms of the molecule, there may be a number of applications for these thermostable variants (such as MTH0895 and MTH0807) in non-medical or industrial fields. In particular, MTH0807 is a very stable molecule – remaining active after more than three years (and counting) at room temperature. Given that thioredoxins have been shown to improve the quality of dough and give many food products enhanced nutritional value along with decreased allergenicity (*9, 10, 11*), this suggests that MTH0807 could be an ideal candidate additive for a variety of baking applications. Specifically, the MTH0807 gene could be inserted into bakers' yeast and the protein produced as the yeast grows. MTH0807 could also be explored in medicine as a first-line antidote to venoms from snakes, scorpions and bees. Many types of antivenom need refrigeration and most often produce severe anaphylactic shock due to their mode of preparation. MTH0807 is stable, does not need refrigeration, and is easily purified in high yields. Thus allergic reactions may be minimal. It could be used as a first aid in the field as the patient waits or is being transported for professional medical aid.

In Chapter 4, I described the determination of the solution structure of another archaebacterial protein called MTH0776. While we were unable to determine a precise function for the protein, the most significant finding was that this protein belongs to a family of proteins with a completely novel fold. These days, only a small percentage (<3% of deposited structures, http://scop.mrc-lmb.cam.ac.uk/scop/count.html#scop-1.67) are found to belong to novel protein fold families. Because relatively little information could be gained from our structural analyses, we turned to computational approaches. Bioinformatic studies have shown that MTH0776 is likely to be expressed together with MTH0777 in a two protein operon system. Furthermore, these two proteins are only found in methanogenic bacteria. This suggests a possible role in methanogenesis. This hypothesis could be investigated further by performing gene knockouts of MTH0776 and MTH0777 to observe their effects on the growth of the archaeabacterium or in the production of methane.

Structural analysis of MTH0777 (the other member of this two-component operon) using its recently determined X-ray structure (1KJN) shows that it may belong to a group of proteins (KTN/RCK domain proteins) involved mainly in transport of small molecules. These proteins normally associate with each other using hydrophobic patches on their surfaces. Some of these hydrophobic patches are absent in MTH0777 (lacks the extra β-α secondary structure at the C-terminus) but MTH0776 has several exposed hydrophobic residues which may serve as binding surfaces for the MTH0777 molecule. Preliminary binding/association studies using a mixture of MTH0776 (^{15}N labelled) and an unlabelled sample of MTH0777 were performed to investigate this possibility. The resulting ^{15}N-HSQC spectrum (figure 4.02 and figure 4.13), size exclusion column chromatography and ^{15}N T_2 relaxation measurements (appendix 2) showed that MTH0776 and MTH0777 do associate as evidenced by the reduced T_2, changes in chemical shifts and resonance intensities of a number of residues found on or near a surface groove. Since these domains (KTN and RCK) are found mainly in uptake or efflux systems (transport systems) we propose that MTH0776 and MTH0777 form complexes that lead to the transport of particular molecules/ligands which is specific to the methanogenic archaebacteria. Detailed experiments are however needed to determine the exact nature of the binding between MTH0776 and MTH0777.

Reference

1. Watson, J.D., Laskowski, R.A. and Thornton, J.M (2005) *Curr. Opin. Struct. Biol.*, **15**, 275 – 284.
2. Laskowski, R.A., Watson, J.D. and Thornton, J.M (2003) *J. Struct. Func. Genomics*, **4**, 167 – 177.
3. Laskowski, R.A., Watson, J.D. and Thornton, J.M (2005) *Nuc. Acids Res.*, **33**, W89 – W93.
4. Turunen, N., Karihtala, P., Mäntyniemi, A., Sormune, R., Holmgren, A., Vuokko, L. and Ylermi, S. (2004) *APMIS*, **112**, 123 – 132.
5. ProLX Pharmaceuticals, Corp., www.prolx.com/content/Thioredoxin.asp
6. Yokomise, H., Fukuse, T., Hirata, T., Go, T., Muro, K., Yagi, K., Unui, K., Hitomi, S., Mitsui, A., Hirakawa, T., Yodoi, J. and Wada, H. (1994) *Respiration*, **61**, 99 – 104.
7. Buchanan, B.B., Schürmann, P. and Jacquot, J-P. (1994) In Seminars in Cell Biology (Grossman, A.R., Ed) Vol **5**, Academic Press London.
8. Yoshida, S., Katoh, T., Tetsuka, T., Uno, K., Matsui, N. and Okamoto, T. (1999), *J. Immunol.*, **163**, 351-358.
9. Buchanan, B.B., Schürmann, P., Decottignies. P. and Lozanos, R.M. (1994) *Arch. Boichem. Biophys.*, **314**, 257 – 260.
10. del Val, G., Yee, B.C., Lozano, R.M., Buchanan, B.B., Ermel, R.E., Lee, Y.M. and Frick, O.L. (1999) *J. Aller. Clin. Immunol.*, **104**, 690 – 697.
11. Buchanan, B.B. (2001) *Plant Physiol.*, **26**, 5 – 7.

Appendix 1

Statistical analysis of the glutaredoxin/thioredoxin activity of T4 Grx, E. coli Grx, E. coli Trx, MTH0807 and MTH0895

Single sample t-test statistical analysis of the glutaredoxin and thioredoxin activities of MTH0807, MTH0895, E. coli thioredoxin, E. coli glutaredoxin and T4 glutaredoxin. The change in absorbance of each variable point at time t (min) as shown in the table is the mean of three readings and SD is their standard deviation from the mean. The p-values (2-sided t-test) for each sample were obtained by comparison to T4 glutaredoxin and E. coli glutaredoxin in the glutaredoxin activity test or T4 glutaredoxin and E. coli thioredoxin in the thioredoxin activity test. A low p-value supports rejection of the null hypothesis i.e. there is a significant difference between the data points compared. The comparisons were made in the first two or three data points after the start of the reaction as the reaction is essentially complete by then.

Table 1A1 Single sample 2-sided t-test analysis of the glutaredoxin activity test of MTH0807, MTH0895, E. coli thioredoxin, E. coli glutaredoxin and T4 glutaredoxin.

Time (Min)	Mean T4 Grx	SD T4 Grx	Mean E.C.Grx	SD E.C. Grx	Mean MTH0807	SD MTH0807	Mean MTH0895	SD MTH0895	Mean E.C. Trx	SD E.C. Trx
1	0.002	0.002	0.001	0.000	0.002	0.001	0.003	0.001	0.002	0.001
2	0.002	0.001	0.003	0.001	0.002	0.000	0.003	0.001	0.001	0.000
3	0.002	0.001	0.002	0.000	0.002	0.001	0.003	0.001	0.003	0.000
4	0.003	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.000
5	0.002	0.001	0.003	0.001	0.002	0.000	0.002	0.001	0.002	0.001
6	0.157	0.004	0.132	0.004	0.061	0.005	0.041	0.003	0.044	0.001
7	0.176	0.004	0.135	0.004	0.065	0.004	0.044	0.005	0.045	0.000
8	0.178	0.004	0.137	0.005	0.066	0.005	0.043	0.004	0.046	0.003
9	0.177	0.004	0.137	0.005	0.069	0.003	0.043	0.004	0.048	0.002
10	0.175	0.004	0.138	0.003	0.069	0.004	0.044	0.004	0.048	0.001
11	0.178	0.005	0.138	0.004	0.069	0.002	0.044	0.003	0.047	0.001
12	0.176	0.004	0.140	0.002	0.070	0.002	0.043	0.004	0.049	0.002
13	0.177	0.003	0.141	0.002	0.070	0.003	0.044	0.004	0.049	0.004
14	0.175	0.002	0.140	0.002	0.071	0.003	0.045	0.004	0.048	0.003
15	0.176	0.004	0.140	0.003	0.069	0.003	0.046	0.005	0.048	0.002
16	0.178	0.004	0.139	0.003	0.070	0.003	0.046	0.004	0.049	0.003
17	0.177	0.003	0.140	0.003	0.070	0.003	0.046	0.004	0.050	0.002
18	0.176	0.003	0.141	0.003	0.071	0.003	0.046	0.005	0.048	0.003
19	0.177	0.003	0.141	0.002	0.070	0.003	0.047	0.004	0.048	0.003
20	0.176	0.001	0.139	0.003	0.071	0.003	0.048	0.004	0.049	0.003
21	0.177	0.004	0.141	0.005	0.070	0.004	0.047	0.004	0.050	0.003
22	0.177	0.003	0.141	0.003	0.071	0.003	0.047	0.003	0.051	0.003
23	0.178	0.003	0.141	0.002	0.070	0.004	0.047	0.004	0.051	0.002
24	0.178	0.005	0.141	0.004	0.070	0.003	0.047	0.004	0.051	0.002
25	0.178	0.003	0.140	0.003	0.071	0.003	0.048	0.003	0.051	0.002

t-test in comparison to T4 Glutaredoxin

Sample	E. coli Grx	E. coli Trx	MTH0807	MTH0895
p-value (6 min)	$p = 0.008$	$p < 0.001$	$p < 0.001$	$p < 0.001$
p-value (7 min)	$p = 0.003$	$p < 0.001$	$p < 0.001$	$p < 0.001$

t-test in comparison to E. coli Glutaredoxin

Sample	T4 Grx	E. coli Trx	MTH0807	MTH0895
p-value (6 min)	$p = 0.008$	$p < 0.001$	$p = 0.002$	$p < 0.001$
p-value (7 min)	$p = 0.003$	$p < 0.001$	$p = 0.001$	$p = 0.001$

Table 1A2 Single sample 2-sided t-test analysis of the thioredoxin activity test of MTH0807, MTH0895, E. coli thioredoxin, and T4 glutaredoxin (OXIDATION OF NADPH).

Time (Min)	Mean T4 Grx	SD T4Grx	Mean E.C. Trx	SD E.C. Trx	Mean MTH0807	SD MTH0807	Mean MTH0895	SD MTH0895
1	0.002	0.001	0.003	0.001	0.003	0.002	0.002	0.001
2	0.002	0.001	0.002	0.001	0.003	0.001	0.002	0.000
3	0.002	0.000	0.002	0.001	0.002	0.001	0.003	0.001
4	0.003	0.002	0.002	0.001	0.003	0.001	0.002	0.001
5	0.305	0.004	0.214	0.018	0.332	0.020	0.185	0.029
6	0.372	0.021	0.245	0.007	0.427	0.026	0.210	0.022
7	0.405	0.011	0.270	0.002	0.666	0.054	0.231	0.013
8	0.424	0.012	0.294	0.001	0.747	0.056	0.270	0.025
9	0.442	0.010	0.322	0.021	0.876	0.035	0.302	0.011
10	0.468	0.004	0.332	0.020	0.966	0.045	0.315	0.002
11	0.497	0.003	0.346	0.022	1.047	0.035	0.330	0.009
12	0.520	0.003	0.367	0.024	1.176	0.025	0.347	0.015
13	0.537	0.004	0.378	0.024	1.222	0.030	0.362	0.021
14	0.563	0.010	0.390	0.022	1.261	0.050	0.369	0.027
15	0.576	0.004	0.414	0.016	1.297	0.066	0.377	0.025
16	0.590	0.004	0.423	0.017	1.347	0.054	0.394	0.020
17	0.607	0.005	0.447	0.010	[illegible]	0.050	0.408	0.014
18	0.595	0.022	0.465	0.013	[illegible]	0.040	0.426	0.003
19	0.609	0.014	0.473	0.010	[illegible]	0.043	0.440	0.011
20	0.617	0.014	0.487	0.014	[illegible].543	0.041	0.450	0.018
21	0.632	0.010	0.507	0.012	1.578	0.044	0.456	0.013
22	0.636	0.015	0.525	0.008	1.618	0.065	0.465	0.012
23	0.658	0.016	0.534	0.002	1.642	0.060	0.472	0.007
24	0.679	0.016	0.558	0.004	1.683	0.080	0.483	0.003
25	0.706	0.014	0.566	0.007	1.722	0.090	0.494	0.007

t-test in comparison to T4 Glutaredoxin

Sample	E. coli Trx	MTH0807	MTH0895
p-value (5 min)	$p = 0.013$	$p = 0.144$	$p = 0.019$
p-value (6 min)	$p = 0.001$	$p = 0.067$	$p = 0.006$
p-value (7 min)	$p < 0.001$	$p = 0.014$	$p = 0.002$

t-test in comparison to E. coli Thioredoxin

Sample	T4 Grx	MTH0807	MTH0895
p-value (5 min)	$p < 0.001$	$p = 0.009$	$p = 0.225$
p-value (6 min)	$p = 0.009$	$p = 0.007$	$p = 0.110$
p-value (7 min)	$p = 0.002$	$p = 0.006$	$p = 0.035$

Table 1A3 Single sample 2-sided t-test analysis of the thioredoxin activity test of MTH0807, MTH0895, E. coli thioredoxin, and T4 glutaredoxin (REDUCTION OF DTNB).

Time (Min)	Mean T4 Grx	SD T4 Grx	Mean E.C. Trx	SD E.C. Trx	Mean MTH0807	SD MTH0807	Mean MTH0 895	SD MTH0895
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	1.073	0.038	0.747	0.004	0.094	0.003	0.138	0.018
3	1.454	0.198	1.303	0.009	0.101	0.001	0.191	0.011
4	1.773	0.119	1.638	0.015	0.110	0.002	0.273	0.020
5	1.976	0.037	1.756	0.005	0.119	0.003	0.362	0.051
6	2.020	0.012	1.784	0.011	0.117	0.005	0.519	0.007
7	2.131	0.030	1.828	0.006	0.127	0.005	0.554	0.022
8	2.111	0.030	1.849	0.008	0.120	0.003	0.596	0.010
9	2.108	0.012	1.856	0.004	0.143	0.019	0.703	0.002
10	2.101	0.023	1.857	0.006	0.116	0.003	0.766	0.047
11	2.096	0.027	1.861	0.011	0.108	0.010	0.873	0.050
12	2.108	0.014	1.866	0.011	0.115	0.003	0.911	0.042
13	2.104	0.015	1.866	0.012	0.121	0.000	0.969	0.043
14	2.154	0.037	1.867	0.009	0.118	0.009	1.021	0.046
15	2.102	0.016	1.862	0.008	0.123	0.004	1.067	0.056
16	2.099	0.021	1.863	0.007	0.130	0.002	1.101	0.056
17	2.102	0.013	1.862	0.010	0.127	0.001	1.146	0.067
18	2.088	0.026	1.863	0.007	0.124	0.008	1.176	0.063
19	2.100	0.018	1.864	0.010	0.133	0.004	1.185	0.068
20	2.106	0.015	1.865	0.009	0.134	0.002	1.202	0.070
21	2.113	0.016	1.867	0.008	0.133	0.005	1.202	0.067
22	2.091	0.027	1.867	0.006	0.130	0.003	1.208	0.067
23	2.089	0.030	1.871	0.006	0.122	0.006	1.214	0.072
24	2.069	0.051	1.868	0.006	0.128	0.001	1.217	0.074
25	2.067	0.048	1.867	0.003	0.125	0.006	1.227	0.084

t-test in comparison to T4 Glutaredoxin

Sample	E. coli Trx	MTH0807	MTH0895
p-value (2 min)	$p < 0.001$	$p < 0.001$	$p < 0.001$
p-value (3 min)	$p = 0.001$	$p < 0.001$	$p < 0.001$
p-value (4 min)	$p = 0.004$	$p < 0.001$	$p < 0.001$

t-test in comparison to E. coli Thioredoxin

Sample	T4 Grx	MTH0807	MTH0807
p-value (2 min)	$p = 0.004$	$p < 0.001$	$p < 0.001$
p-value (3 min)	$p = 0.317$	$p < 0.001$	$p < 0.001$
p-value (4 min)	$p = 0.188$	$p < 0.001$	$p < 0.001$

Appendix 2

Measurement of ^{15}N T_2 relaxation of MTH0776 and MTH0776 in complex with MTH0777

Table 2A1. Table of ^{15}N T_2 relaxation times for 0.5 mM solution of MTH0776 in buffer solution consisting of 50 mM NaH_2PO_4, 300 mM NaCl, 15mM DTT and 1.0 mM DSS.

Residue	T_2(msec)	SD (msec)	Residue	T_2(msec)	SD (msec)
1M	138.33	3.74	49G	114.57	4.01
2T	105.56	3.48	50A	110.92	3.00
3F	120.78	1.93	52I	121.82	5.97
5L	105.99	4.13	53I	114.16	1.94
6E	103.05	1.75	54G	105.34	2.32
8Y	136.21	2.18	55I	150.41	18.65
9L	102.54	2.46	58V	109.40	0.77
10Q	99.35	1.99	60I	105.14	2.10
11Q	94.22	1.88	61G	108.45	5.53
12S	107.52	4.19	62I	110.39	2.65
13G	110.02	2.75	63D	105.12	3.15
14E	103.52	1.76	65E	115.59	1.04
15Y	97.82	1.57	66R	104.06	2.50
16E	92.14	3.23	67S	99.07	1.59
17I	100.2	2.31	68T	98.72	3.06
18H	99.85	2.20	69V	103.01	2.78
19M	97.59	1.95	70M	113.96	2.28
20K	119.68	3.47	71C	105.82	1.16
21R	109.73	1.87	73Y	134.03	11.26
22A	104.85	2.83	74T	108.92	6.21
23G	111.65	4.13	75K	108.28	6.17
24F	106.38	1.60	77C	109.33	3.50
25R	93.11	3.72	78Y	106.05	3.50
26E	105.75	1.80	79G	116.37	3.26
27C	57.63	0.92	80T	101.57	2.24
28A	97.66	2.05	81A	110.84	5.76
29A	97.54	2.34	81R	101.04	3.33
30M	92.49	1.02	82V	98.07	1.86
31I	99.62	2.39	83V	110.20	3.31
32E	102.18	3.27	84E	113.34	3.39
33K	97.61	1.56	85L	114.27	3.66
34K	94.76	1.61	87V	102.85	3.81
35A	123.61	3.59	88D	118.76	2.02
36R	112.22	2.02	91E	104.47	2.93
37R	101.31	2.33	92I	93.19	3.54
38V	108.79	3.16	93E	103.16	3.10
39V	106.78	2.88	94R	99.20	2.08
40H	96.04	1.54	95I	104.68	3.14
41I	110.87	5.32	96L	101.10	1.62
42K	111.27	4.79	97E	106.11	2.87
44G	112.15	4.15	98V	59.19	4.14
45E	102.35	2.35	99A	116.88	3.16
46K	109.87	1.87	100E	116.69	1.75
47I	102.61	2.77			
48L	119.19	4.41			

Table 2A2. Table of ^{15}N T_2 relaxation times for 0.5 mM solution of MTH0776 in the presence of 1.0 mM of MTH0777 in a buffer solution consisting of 50 mM NaH_2PO_4, 300 mM NaCl, 15mM DTT and 1.0 mM DSS.

Residue	T_2 msec	SD (msec)
2T	73.67	3.32
3F	84.60	3.05
6E	68.15	7.16
9L	72.28	2.53
10Q	49.84	1.95
11Q	50.74	3.60
13G	72.79	7.21
14E	67.35	1.68
15Y	56.03	2.02
19M	65.57	5.11
20K	112.23	4.49
23G	56.40	4.17
24F	98.49	3.25
27C	49.85	3.64
30M	51.86	5.08
32E	91.36	3.02
34E	102.73	4.32
34K	71.77	2.37
35A	113.25	4.42
36R	50.12	13.18
37R	49.95	1.19
39V	67.11	7.18
45E	68.23	2.59
46K	93.77	3.28
49G	42.05	4.37
52I	71.43	11.43
62I	36.52	4.49
63D	84.96	1.19
67S	68.03	6.06
68T	71.46	7.72
75K	80.54	16.91
77C	45.46	7.09
80T	68.13	19.21
81R	69.44	5.90
82V	72.09	1.73
83V	53.65	3.82
87V	71.38	8.42
88D	66.11	5.02
93E	50.76	1.02
94R	49.87	1.30
95I	73.51	3.38
97E	52.27	2.72
98V	22.23	1.96
99A	51.87	2.65
100E	45.30	7.70

Table 2A3 MTH0807 resonance assignments in 500 μL of a buffer solution (pH 6.0) made up of 50 mM NaH_2PO_4, 100 mM NaCl, 50 μL of D_2O, 1 mM DSS and 10 μL of a 3% solution of sodium azide. Proton chemical shifts are given in parentheses.

Residue	N	C	C^{α}	C^{β}	other
M1	121.29 (8.42)	175.51	55.78 (4.47)	33.04 (2.01, 2.01)	C^{γ}, * (2.54, 2.54)
V2	120.91 (8.02)	175.27	61.32 (4.17)	34.26 (1.80)	$C^{\gamma1}$, * (0.86); $C^{\gamma2}$, * (0.70)
V3	126.31 (8.17)	174.66	62.41 (4.18)	32.84 (1.97)	$C^{\gamma1}$, * (0.90); $C^{\gamma2}$, * (0.69)
N4	127.08 (8.65)	174.39	53.88 (5.06)	40.50 (2.78, 2.38)	
I5	127.08 (9.34)	174.82	59.63 (5.05)	39.32 (1.95)	$C^{\gamma1}$, * (1.24, 1.10); $C^{\gamma2}$, * (0.80); $C^{\delta1}$, * (0.73)
E6	127.46 (9.06)	174.35	54.29 (5.04)	33.91 (2.06, 2.06)	C^{γ}, * (1.45, 1.45)
V7	122.45 (8.80)	175.50	59.64 (4.94)	33.90 (1.87)	$C^{\gamma1}$, * (0.84); $C^{\gamma2}$, * (0.84)
F8	126.69 (9.40)	174.04	57.34 (5.42)	40.94 (2.95, 2.81)	$C^{\delta1}$, * (6.80); $C^{\delta2}$, * (6.80); $C^{\epsilon1}$, * (7.05); $C^{\epsilon2}$, * (7.05); C^{ζ}, * (6.87)
T9	110.88 (7.99)	172.01	59.58 (4.37)	68.44 (4.75)	
S10	113.58 (7.61)		56.78 (5.12)	65.25 (3.71, 3.46)	
P11		176.30	63.87 (4.80)	(2.78, 1.93)	C^{γ}, * (1.65, 1.65); C^{δ}, * (3.28, 3.14)
T12	115.12 (8.23)	177.10	58.44 (4.41)	63.87 (3.82)	$C^{\gamma2}$, * (1.62)
C13	124.76 (8.10)		57.02 (4.31)	30.02 (3.87, 3.87)	
P14		174.18			
Y15	112.04 (9.25)	177.89	59.58 (4.99)	40.91 (2.99, 2.99)	$C^{\delta1}$, * (7.05); $C^{\delta2}$, * (7.05); $C^{\epsilon1}$, * (6.87); $C^{\epsilon2}$, * (6.87)
C16	116.67 (7.47)		56.09 (4.27)	30.00 (2.43, 2.15)	
P17		176.70	66.68 (4.52)	(1.97, 1.97)	C^{γ}, * (1.27, 1.27); C^{δ}, * (3.90, 3.77)
M18	115.89 (7.54)	178.45	59.27 (4.29)	32.39 (2.37, 2.16)	C^{γ}, * (2.71, 2.65); C^{ϵ}, * (1.14)
A19	122.06 (7.72)	177.88	54.89 (4.10)	18.11 (1.40)	
I20	116.67 (8.10)	179.42	63.85 (3.49)	36.74 (1.97)	$C^{\gamma1}$, * (0.84, 0.84); $C^{\gamma2}$, * (0.74)
E21	118.59 (8.08)	179.51	59.91 (4.12)	29.36 (2.16, 2.16)	C^{γ}, * (2.43, 2.30)
V22	116.28 (7.65)	177.80	65.56 (4.03)	31.77 (2.11)	$C^{\gamma1}$, * (1.03); $C^{\gamma2}$, * (1.03)
V23	122.06 (7.75)	177.88	66.76 (3.71)	30.93 (2.21)	$C^{\gamma1}$, * (0.79); $C^{\gamma2}$, * (0.89)
D24	123.22 (8.82)	180.61	57.64 (4.52)	39.35 (2.93, 2.57)	
E25	121.68 (8.26)	179.93	59.88 (4.15)	28.78 (2.26, 1.59)	C^{γ}, * (2.56, 2.56)
A26	123.22 (8.38)	179.00	55.01 (4.10)	16.79 (1.51)	
K27	120.91 (8.66)	175.38	58.35 (4.09)	31.50 (1.92, 1.68)	C^{γ}, * (1.10, 1.10); C^{δ}, * (1.42, 1.42); C^{ϵ}, * (2.29, 2.29); N^{ζ}, * (7.27)
K28	120.14 (7.81)	177.87	59.44 (4.02)	32.30 (1.73, 1.58)	C^{γ}, * (1.27, 1.27); C^{δ}, * (1.42, 1.42); C^{ϵ}, * (2.00, 2.00); N^{ζ}, * (7.52)
E29	118.21 (7.61)	177.72	58.29 (4.01)	30.14 (1.42, 1.26)	C^{γ}, * (1.74, 1.74)
F30	113.58 (8.77)	173.58	59.39 (4.38)	40.74 (2.98, 2.63)	$C^{\delta1}$, * (6.69); $C^{\delta2}$, * (6.69); $C^{\epsilon1}$, * (7.20); $C^{\epsilon2}$, * (7.20); C^{ζ}, * (6.97)
G31	111.63 (8.30)	175.63	47.39 (3.93, 3.31)		
D32	125.53 (8.85)	176.73	55.29 (4.51)	39.98 (2.80, 2.35)	
K33	117.44 (7.92)	176.56	57.98 (4.28)	33.39 (2.11, 1.91)	C^{γ}, * (1.52, 1.52); C^{δ}, * (1.72, 1.73); C^{ϵ}, * (2.58, 2.58); N^{ζ}, * (7.49)
I34	108.18 (7.60)	176.87	58.27 (5.34)	41.51 (1.68)	$C^{\gamma1}$, * (1.80, 1.80); $C^{\delta1}$, * (0.48)
D35	122.06 (8.52)	174.43	52.90 (4.95)	44.12 (2.69, 2.37)	

V36	123.99 (8.55)	174.98	61.32 (4.76)	33.96 (1.91)	$C^{\gamma1}$, * (0.86); $C^{\gamma2}$, * (0.70)
E37	129.01 (8.91)	174.13	54.76 (4.60)	32.48 (1.91, 1.80)	C^{γ}, * (2.21, 2.13)
K38	125.92 (8.63)	175.57	55.42 (4.79)	32.73 (1.72, 1.61)	C^{γ}, * (0.95, 0.95); C^{δ}, * (1.24, 1.24); C^{ϵ}, * (3.28, 3.14); N^{ζ}, * (7.75)
I39	127.85 (9.17)	173.76	60.90 (4.14)	40.94 (1.25)	$C^{\gamma1}$, * (0.93, 0.93); $C^{\delta1}$, * (0.87)
D40	128.62 (8.84)	178.69	52.11 (4.99)	41.80 (2.86, 2.34)	
I41	121.29 (9.02)	176.68	63.60 (3.86)	38.15 (1.87)	$C^{\gamma1}$, * (1.68, 1.14); $C^{\gamma2}$, * (0.79); $C^{\delta1}$, * (0.63)
M42	118.59 (9.21)	177.82	55.61 (4.57)	30.58 (2.54, 2.25)	C^{γ}, * (2.78, 2.78); C^{ϵ}, * (2.14)
V43	118.98 (7.71)	176.01	63.75 (4.05)	33.51 (2.18)	$C^{\gamma1}$, * (0.96); $C^{\gamma2}$, * (0.96)
D44	121.68 (8.28)	175.13	53.26 (4.97)	41.61 (2.84, 2.45)	
R45	121.29 (8.19)	178.08	57.15 (4.34)	31.07 (1.96, 1.83)	C^{γ}, * (1.64, 1.64); C^{δ}, * (3.24, 3.24)
E46	118.21 (8.44)	179.23	60.57 (3.94)	29.22 (2.08, 1.95)	C^{γ}, * (2.36, 2.12)
K47	119.36 (8.38)	177.70	58.17 (3.99)	31.91 (1.90, 1.65)	C^{γ}, * (1.02, 1.02); C^{δ}, * (1.38, 1.38); N^{ζ}, * (7.50)
A48	120.52 (7.91)	179.95	55.19 (3.62)	17.83 (1.41)	
I49	117.82 (7.99)	175.13	64.28 (3.80)	37.91 (1.92)	$C^{\gamma1}$, * (0.86, 0.86)
E50	122.84 (8.00)	178.83	59.44 (3.93)	29.27 (2.04, 1.90)	C^{γ}, * (2.26, 2.26)
Y51	115.12 (7.97)	176.68	56.80 (4.61)	38.43 (2.97, 2.54)	$C^{\delta1}$, * (6.81); $C^{\delta2}$, * (6.81)
G52	110.50 (7.93)	175.20	46.47 (3.29, 3.85)		
L53	117.44 (7.79)	176.84	55.50 (4.01)	42.64 (1.50, 1.50)	C^{γ}, * (1.45); $C^{\delta1}$, * (1.25); $C^{\delta2}$, * (1.25)
M54	122.06 (8.38)	174.87	59.61 (4.59)	32.80 (2.47, 2.47)	C^{γ}, * (1.98, *)
A55	121.68 (7.57)	174.38	51.49 (4.55)	21.91 (1.25)	
V56	107.41 (7.69)		57.95 (4.43)	33.87 (2.30)	$C^{\gamma1}$, * (0.93); $C^{\gamma2}$, * (0.79)
P57		174.77	61.57 (4.60)	35.24 (1.91, 1.91)	C^{γ}, * (1.80, *); C^{δ}, * (3.14, 2.99)
A58	119.75 (8.21)	175.83	51.26 (5.39)	23.87 (1.61)	
I59	120.14 (9.42)	176.11	60.19 (5.44)	42.64 (1.98)	$C^{\gamma1}$, * (1.05, 1.05); $C^{\delta1}$, * (0.95)
A60	131.63 (9.95)	175.87	50.01 (5.41)	21.78 (1.28)	
I61	120.52 (8.66)	171.97	59.99 (4.09)	39.89 (1.71)	$C^{\gamma1}$, * (1.45, 1.14); $C^{\gamma2}$, * (0.80); $C^{\delta1}$, * (0.65)
N62	127.85 (9.98)	172.78	53.92 (4.63)	37.06 (3.21, 2.64)	$N^{\delta2}$, * (7.60, 7.16)
G63	127.46 (9.39)	173.65	45.52 (4.00, 4.05)		
V64	118.59 (7.33)	176.15	60.67 (4.12)	35.05 (1.90)	$C^{\gamma1}$, * (0.88); $C^{\gamma2}$, * (0.79)
V65	127.85 (9.03)	175.15	65.60 (4.31)	30.66 (1.89)	$C^{\gamma1}$, * (0.84); $C^{\gamma2}$, * (0.80)
R66	128.23 (8.69)	175.02	53.96 (4.34)	32.53 (1.25, 1.25)	C^{γ}, * (1.12, 1.12); N^{ϵ}, * (7.17)
F67	114.35 (7.76)	174.97	56.82 (5.04)	42.69 (3.23, 2.58)	$C^{\delta1}$, * (7.04); $C^{\delta2}$, * (7.04); $C^{\epsilon1}$, * (7.18); $C^{\epsilon2}$, * (7.18); C^{ζ}, * (7.04)
V68	123.61 (8.69)	176.20	62.87 (4.49)	32.25 (1.97)	$C^{\gamma1}$, * (0.83); $C^{\gamma2}$, * (0.76)
G69	114.35 (8.21)	174.35	43.75 (4.32, 3.82)		
A70	121.29 (8.78)		49.94 (4.48)	22.67 (1.42)	
P71		175.56	61.49 (4.00)	31.31 (1.90, 1.90)	C^{γ}, * (1.47, 1.47); C^{δ}, * (3.77, 3.24)
S72	113.97 (8.61)	174.98	57.05 (4.36)	65.13 (3.96, 3.57)	
R73	122.45 (8.86)	177.06	59.84 (3.45)	29.80 (1.91, 1.91)	C^{γ}, * (1.37, 1.62); C^{δ}, * (3.39, 3.33)
E74	115.89 (8.61)	179.57	60.62 (3.75)	28.81 (1.92, 1.92)	
E75	118.59 (7.75)	179.83	58.89 (4.13)	30.02 (2.13, 1.92)	C^{γ}, * (2.38, 2.24)
L76	120.52 (8.15)	177.59	57.91 (4.13)	40.88 (1.27, 1.27)	C^{γ}, * (1.21); $C^{\delta1}$, * (0.95); $C^{\delta2}$, * (0.87)
F77	119.75 (9.40)	178.74	60.49 (4.34)	36.65 (3.50, 3.29)	$C^{\delta1}$, * (6.70); $C^{\delta2}$, * (6.70); $C^{\epsilon1}$, * (7.17); $C^{\epsilon2}$, * (7.17);

					C^{ϵ}, * (6.70)
E78	118.98 (8.11)	177.78	59.75 (4.19)	29.46 (2.18, 1.60)	C^{γ}, * (2.50, 2.30)
A79	120.14 (7.64)	180.34	54.89 (4.31)	18.32 (1.67)	
I80	117.82 (8.11)	177.56	66.23 (3.28)	37.93 (1.45)	$C^{\gamma 2}$, * (0.46); $C^{\delta 1}$, * (-0.27)
N81	117.82 (8.45)	177.23	56.91 (4.37)	38.79 (2.96, 2.69)	
D82	120.91 (8.06)	176.58	51.73 (4.41)	42.79 (2.05, 2.05)	
E83	124.38 (8.64)	174.25	59.93 (4.50)	32.54 (1.92, 1.92)	C^{γ}, * (2.08, 2.08)
M84	125.15 (7.74)	176.85	57.02 (4.05)	29.88 (1.99, 1.88)	C^{γ}, * (2.22, *)
E85	118.21 (7.51)		56.01 (4.38)	33.91 (2.13, 2.13)	C^{γ}, * (2.67, 2.52)

Table 2A4 MTH0776 resonance assignments in 500 μL of a buffer (pH 6.8) made up of 20 mM NaH_2PO_4, 300 mM NaCl, 50 μL of D_2O, 1 mM 2, 2 -dimethyl-2-silapentane-5-sulfonic acid (DSS), 15 mM DTT and 10 μL of a 3% solution of sodium azide at 25°C. Proton chemical shifts are given in parentheses. First 20 amino acid belong to the Histidine tag.

Residue	N	C	C^{α}	C^{β}	other
M1		177.21	56.18 (4.03)	31.22 (1.78, 1.78)	C^{γ}, * (1.94, 1.94)
G2	114.85 (7.98)	174.33	45.35 (4.05, 4.05)		
S3	115.33 (8.17)	175.56	58.04 (4.51)	63.91 (3.76, 3.76)	
S4	115.03 (8.07)		58.34 (4.05)	64.86 (3.24, 3.24)	
H6		175.58	(4.35)	(3.18, 3.18)	
H7	122.08 (8.21)	176.54	56.57 (4.44)	30.54 (3.12, 3.00)	
H8	120.63 (8.10)	177.52	56.96 (4.51)	30.53 (2.86, 2.86)	
H9	121.11 (8.21)	176.13	56.49 (4.37)	30.38 (3.24, *)	
H10	124.49 (8.33)	176.00	55.94 (4.66)	30.80 (2.73, 2.73)	
S11	121.11 (8.72)	174.83	58.48 (4.46)	63.88 (4.04, 4.04)	
S12	117.74 (8.46)	176.97	58.64 (4.54)	63.81 (3.97, 3.97)	
G13	110.03 (8.40)	173.84	45.29 (4.03, 4.03)		
L14	121.11 (8.03)	177.06	55.07 (4.42)	42.39 (1.87, 1.67)	C^{γ}, 26.93 (1.34); $C^{\delta1}$, 24.88 (1.16); $C^{\delta2}$, 23.36 (1.16)
V15	121.60 (8.07)		59.61 (4.10)	32.67 (1.96)	$C^{\gamma1}$, * (0.95); $C^{\gamma2}$, * (0.85)
P16		176.12	63.10 (4.45)	31.98 (2.14, 2.14)	C^{γ}, 24.06 (1.96, 1.69); C^{δ}, 51.61 (3.24, 3.24)
R17	122.56 (8.45)	177.12	56.31 (4.40)	30.78 (1.74, 1.74)	C^{γ}, 27.07 (1.61, 1.61); C^{δ}, 43.34 (3.24, 3.00)
G18	110.03 (8.51)	174.33	45.38 (4.03, 4.03)		
S19	115.81 (8.24)	175.86	58.52 (4.70)	63.89 (3.65, 3.65)	
H20	123.04 (8.81)	174.45	56.83 (4.51)	33.09 (2.93, 2.93)	
M21	121.11 (8.23)	175.89	55.62 (4.30)	32.97 (2.24, 2.24)	C^{γ}, 31.98 (2.64, 2.49)
T22	114.85 (8.13)	173.90	61.67 (4.41)	69.84 (3.98)	$C^{\gamma2}$, 21.46 (1.25)
F23	124.49 (8.38)	174.23	58.77 (4.33)	40.22 (2.85, 2.75)	$C^{\delta1}$, 132.21 (7.16); $C^{\delta2}$, 132.21 (7.16); $C^{\epsilon1}$, 131.83 (7.41); $C^{\epsilon2}$, 131.83 (7.41); C^{ζ}, 133.34 (7.25)
C24	126.42 (7.78)	174.48	57.19 (4.57)	28.75 (3.23, 2.86)	
L25	129.31 (8.87)	175.72	58.43 (4.52)	41.84 (2.02, 1.96)	C^{γ}, 25.82 (1.73); $C^{\delta1}$, 24.17 (0.93); $C^{\delta2}$, 23.37 (0.79)
E26	116.78 (8.51)	178.74	60.32 (4.65)	29.34 (2.07, 2.03)	C^{γ}, 32.88 (2.35, 2.35)
T27	115.33 (7.67)	176.12	65.73 (4.21)	68.52 (4.14)	$C^{\gamma2}$, 21.95 (1.22)
Y28	123.52 (8.07)	177.79	62.05 (4.05)	39.02 (3.06, 2.87)	$C^{\delta1}$, 133.32 (7.24); $C^{\delta2}$, 130.83 (7.15); $C^{\epsilon1}$, 119.59 (6.81); $C^{\epsilon2}$, 118.20 (6.72)
L29	115.33 (8.29)	179.20	56.75 (3.94)	41.06 (1.86, 1.25)	C^{γ}, 33.31 (1.20); $C^{\delta1}$, 25.75 (0.91); $C^{\delta2}$, 21.18 (0.69)
Q30	116.29 (7.56)	177.27	57.45 (4.28)	29.27 (2.31, 2.25)	C^{γ}, 33.98 (2.87, 2.54); C^{δ}, 180.37 $N^{\epsilon2}$, 110.73 (7.44, 6.80)
Q31	115.81 (7.85)	175.97	56.09 (4.37)	29.68 (1.98, 1.95)	C^{γ}, 33.77 (2.34, 2.05); C^{δ}, 183.48 $N^{\epsilon2}$, 111.12 (7.51, 6.88)
S32	114.85 (8.22)	174.79	58.31 (4.36)	63.70 (3.62, 3.47)	
G33	110.03 (8.16)	174.45	44.94 (4.18, 4.18)		
E34	120.63 (8.29)	179.22	55.97 (4.46)	29.12 (2.05, 1.64)	C^{γ}, 36.26 (3.09, 2.78)
Y35	116.29 (7.39)	173.31	56.54 (5.02)	39.23 (3.42, 3.07)	$C^{\delta1}$, 133.15 (7.22); $C^{\delta2}$, 133.15 (7.22); $C^{\epsilon1}$, 119.42 (6.87); $C^{\epsilon2}$, 119.42 (6.87)

E36	119.19 (9.07)	175.00	54.93 (4.67)	33.38 (1.93, 1.74)	C^{γ}, 35.92 (2.48, 2.10)
I37	124.01 (8.78)	176.22	59.88 (4.80)	38.25 (1.70)	$C^{\gamma1}$, * (1.46, 1.17); $C^{\gamma2}$, 17.92 (0.91); $C^{\delta1}$, * (0.72)
H38	125.93 (8.62)	177.32	57.49 (4.29)	28.87 (2.86, 2.64)	$C^{\delta2}$, 120.37 (7.16); $C^{\epsilon1}$, 136.86 (8.53)
M39	116.78 (7.85)	175.88	56.13 (4.35)	29.79 (1.86, 1.86)	C^{γ}, 32.16 (2.25, 2.25); C^{ϵ}, 19.23(*)
K40	121.60 (8.36)	177.58	55.30 (4.10)	31.08 (2.07, 2.07)	C^{γ}, 24.89 (1.45, 1.45); C^{δ}, 26.97 (1.89, 1.89); C^{ϵ}, 42.41 (3.43, 3.43)
R41	119.67 (8.06)	176.06	55.10 (4.64)	33.06 (1.93, 1.72)	C^{γ}, 27.27 (1.46, 1.46); C^{δ}, 44.26 (3.25, 3.06)
A42	124.49 (8.95)	177.31	50.57 (4.89)	23.04 (1.40)	
G43	105.69 (8.54)	174.20	44.83 (4.44, 4.07)		
F44	120.15 (8.27)	175.95	63.71 (4.72)	39.71 (2.79, 2.63)	$C^{\delta1}$, 132.46 (6.97); $C^{\delta2}$, 132.46 (6.97); $C^{\epsilon1}$, 132.30 (7.34); $C^{\epsilon2}$, 132.30 (7.34); C^{ζ}, 130.85 (6.98)
R45	116.78 (8.98)	179.08	59.37 (4.07)	29.03 (2.73, 2.73)	C^{γ}, 27.05 (1.93, 1.31); C^{δ}, 43.53 (3.90, 3.36)
E46	119.67 (8.42)	179.55	59.85 (4.06)	28.81 (2.35, 2.35)	C^{γ}, 37.18 (2.53, 2.53)
C47	119.67 (7.43)	176.13	62.98 (4.16)	27.30 (2.80, 2.80)	
A48	117.26 (8.54)	178.64	55.03 (4.01)	17.89 (1.26)	
A49	117.74 (8.00)	179.08	54.62 (4.08)	17.91 (1.47)	
M50	117.74 (7.37)	177.80	59.04 (4.13)	32.29 (2.02, 1.97)	C^{γ}, 36.31 (2.45, 2.19)
I51	120.15 (7.78)	177.23	66.18 (4.04)	37.84 (1.93)	$C^{\gamma1}$, * (1.89, 1.81); $C^{\gamma2}$, 16.86 (1.01); $C^{\delta1}$, * (0.72)
E52	114.37 (7.98)	178.57	59.28 (3.98)	28.69 (2.29, 2.29)	C^{γ}, 35.98 (2.52, 2.53)
K53	116.29 (7.47)	178.61	57.66 (4.35)	33.65 (1.94, 1.94)	C^{γ}, 24.69 (1.51, 1.29); C^{δ}, 29.03 (1.72, 1.72); C^{ϵ}, * (3.03, 3.03)
K54	116.78 (8.36)	177.52	57.19 (4.23)	34.65 (2.79, 2.54)	C^{γ}, 25.49 (1.59, 1.34); C^{δ}, 28.76 (2.31, 2.31); C^{ϵ}, 41.59 (3.52, 3.52)
A55	120.63 (8.24)	178.20	52.23 (4.36)	20.94 (1.17)	
R56	121.60 (9.00)	175.72	58.73 (4.08)	29.49 (2.05, 2.05)	C^{γ}, 27.20 (1.81, 1.81); C^{δ}, 43.11 (3.32, 3.32)
R57	116.78 (7.57)	174.95	54.98 (4.60)	33.11 (1.97, 1.77)	C^{γ}, 27.23 (1.45, 1.39); C^{δ}, 44.17 (3.71, 3.71)
V58	125.45 (8.83)	175.93	61.06 (4.86)	34.33 (1.80)	$C^{\gamma1}$, 23.32 (0.89); $C^{\gamma2}$, 21.69 (0.62)
V59	120.63 (8.66)	173.66	58.96 (4.66)	33.91 (2.26)	$C^{\gamma1}$, 21.32 (0.83); $C^{\gamma2}$, 18.24 (0.62)
H60	117.26 (8.87)	174.20	54.91 (5.87)	31.71 (3.24, 3.02)	$C^{\delta2}$, 119.61 (7.22); $C^{\epsilon1}$, 137.37 (8.36)
I61	113.88 (8.55)	175.40	58.71 (4.98)	41.76 (2.10)	$C^{\gamma1}$, * (1.34, 1.34); $C^{\gamma2}$, 17.40 (0.85); $C^{\delta1}$, * (0.80)
K62	123.52 (9.03)		54.50 (4.56)	32.07 (2.02, 2.02)	C^{γ}, * (1.57, 1.57); C^{δ}, * (1.74, 1.74); C^{ϵ}, * (2.32, 2.32)
P63		178.01	63.62 (4.22)	31.42 (2.26, 2.04)	C^{γ}, * (1.76, 1.58); C^{δ}, * (3.22, 3.11)
G64	112.92 (9.15)	173.74	45.00 (4.32, 4.32)		
E65	121.60 (7.81)	175.12	56.48 (4.15)	31.06 (2.14, 1.99)	C^{γ}, 32.70 (2.35, 2.35)
K66	123.04 (8.38)	176.27	55.64 (4.91)	33.66 (1.67, 1.67)	C^{γ}, 25.54 (1.12, 1.12); C^{δ}, 29.56 (1.52, 1.37); C^{ϵ}, 43.63 (2.99, 2.99)
I67	124.49 (8.81)	174.75	58.30 (4.37)	38.93 (1.93)	$C^{\gamma1}$, * (1.72, 1.72); $C^{\gamma2}$, 16.99 (0.88); $C^{\delta1}$, * (0.76)
L68	122.56 (9.46)	176.82	55.51 (3.89)	39.00 (1.99, 1.99)	C^{γ}, 26.86 (1.68); $C^{\delta1}$, 25.87 (0.80); $C^{\delta2}$, 22.76 (0.66)
G69	101.51 (8.41)	173.93	45.87 (4.36, 3.64)		
A70	124.97 (8.24)	175.86	50.46 (4.84)	21.78 (1.55)	
R71	120.63 (8.61)	176.22	56.56 (4.29)	30.11 (1.79, 1.71)	C^{γ}, 28.08 (1.50, 1.50); C^{δ}, 43.49 (3.28, 3.28)
I72	127.86 (8.56)	176.65	58.74 (4.36)	34.64 (1.99)	$C^{\gamma1}$, * (1.52, 1.31); $C^{\gamma2}$, 17.55 (1.07); $C^{\delta1}$, * (0.79)
I73	123.52 (7.57)	174.32	60.24 (4.28)	39.29 (1.89)	$C^{\gamma1}$, * (1.34, 1.34); $C^{\gamma2}$, 17.64 (1.13); $C^{\delta1}$, 13.27 (0.86)
G74	105.69 (7.37)	170.50	44.59 (2.99, 2.99)		

I75	120.63 (5.38)		57.23 (4.30)	40.06 (1.95)	$C^{\gamma1}$, * (1.62, 1.36); $C^{\gamma2}$, * (0.99); $C^{\delta1}$, * (0.85)
P76			(4.34)	(2.31, 2.21)	C^{γ}, * (2.04, 2.04); C^{δ}, * (3.06, 3.06)
P77		174.29	59.17 (4.66)	36.87 (2.40, 2.40)	C^{γ}, 24.60 (1.69, 1.69); C^{δ}, 49.54 (3.62, 3.25)
V78	119.69 (8.10)		61.86 (4.42)	33.19 (1.38)	$C^{\gamma1}$, * (0.89); $C^{\gamma2}$, * (0.83)
P79		177.04	61.67 (4.85)	32.40 (2.28, 2.04)	C^{γ}, 25.82 (1.94, 1.74); C^{δ}, 49.17 (3.93, 3.93)
I80	116.29 (8.58)	176.14	57.57 (4.68)	38.69 (1.86)	$C^{\gamma1}$, * (1.56, 1.56); $C^{\gamma2}$, 17.00 (0.93); $C^{\delta1}$, * (0.84)
G81	111.96 (9.46)	172.90	44.77 (4.99, 4.99)		
I82	123.52 (9.35)	174.44	59.57 (4.76)	41.67 (1.70)	$C^{\gamma1}$, * (1.45, 1.17); $C^{\gamma2}$, 17.90 (0.91); $C^{\delta1}$, * (0.76)
D83	125.93 (8.52)	177.44	52.42 (4.97)	41.27 (3.45, 2.69)	
E84	123.52 (9.40)	177.30	58.81 (4.09)	30.21 (2.03, 2.03)	C^{γ}, 38.12 (2.28, 2.28)
E85	120.15 (8.42)	178.50	58.77 (4.24)	29.79 (1.96, 1.72)	C^{γ}, 36.44 (2.46, 2.32)
R86	117.74 (7.88)	175.49	56.19 (4.24)	31.16 (1.89, 1.89)	C^{γ}, 28.33 (1.56, 1.52); C^{δ}, 43.34 (3.24, 3.24)
S87	115.81 (7.95)	173.44	58.91 (3.92)	62.24 (3.66, 3.66)	
T88	107.62 (8.05)	174.23	58.33 (5.48)	72.10 (4.11)	$C^{\gamma2}$, 21.51 (0.86)
V89	115.33 (8.64)	173.40	58.92 (5.21)	35.38 (1.84)	$C^{\gamma1}$, 21.45 (0.76); $C^{\gamma2}$, 21.25 (0.55)
M90	125.93 (9.42)	175.01	54.09 (5.61)	37.16 (2.04, 2.04)	C^{γ}, 31.77 (2.34, 2.34)
I91	120.63 (9.17)		57.25 (5.16)	41.97 (1.98)	$C^{\gamma1}$, * (1.69, 1.69); $C^{\gamma2}$, * (1.05); $C^{\delta1}$, * (0.85)
P92		174.14	60.45 (5.54)	28.56 (2.33, 2.33)	C^{γ}, * (1.42, 1.42); C^{δ}, * (3.54, 3.54)
Y93	128.34 (9.45)	174.31	56.66 (4.87)	41.04 (2.86, 2.69)	$C^{\delta1}$, 131.35 (7.10); $C^{\delta2}$, 131.35 (7.10); $C^{\epsilon1}$, 119.59 (6.93); $C^{\epsilon2}$, 118.35 (6.89)
T94	127.38 (8.22)	175.58	62.61 (4.93)	69.83 (3.93)	$C^{\gamma2}$, 21.07 (1.05)
K95	127.38 (8.39)		52.56 (4.86)	35.17 (1.62, 1.62)	C^{γ}, * (1.33, 1.33); C^{δ}, * (1.46, 1.46); C^{ϵ}, * (2.93, 2.93)
P96		176.06	61.83 (4.66)	33.67 (2.00, 2.00)	C^{γ}, * (1.76, 1.58); C^{δ}, * (3.65, 3.65)
C97	127.38 (9.17)	177.42	61.86 (4.62)	31.55 (2.98, 2.57)	
Y98	118.70 (8.18)	175.12	59.07 (4.13)	40.50 (3.19, 2.87)	$C^{\delta1}$, 130.43 (7.14); $C^{\delta2}$, 130.43 (7.14); $C^{\epsilon1}$, 116.75 (6.86); $C^{\epsilon2}$, 116.75 (6.86)
G99	108.10 (8.23)	172.49	45.18 (4.29, 3.93)		
T100	120.63 (9.18)	173.38	62.73 (5.09)	70.35 (3.85)	$C^{\gamma2}$, 23.05 (1.10)
A101	126.42 (8.74)	177.40	49.59 (4.60)	23.35 (1.49)	
V102	118.70 (8.50)	175.19	62.26 (4.87)	32.84 (1.87)	$C^{\gamma1}$, 21.97 (0.86); $C^{\gamma2}$, * (0.79)
V103	124.97 (8.77)	174.15	59.29 (5.30)	34.83 (1.99)	$C^{\gamma1}$, 21.56 (0.96); $C^{\gamma2}$, 20.03 (0.82)
E104	126.90 (9.43)	174.85	53.79 (4.66)	33.84 (1.76, 1.76)	C^{γ}, 36.07 (2.00, 2.00)
L105	126.42 (9.24)		52.62 (5.32)	43.38 (1.66, 1.66)	C^{γ}, * (1.16); $C^{\delta1}$, * (0.70); $C^{\delta2}$, * (0.70)
P106		176.27	62.15 (5.04)	31.83 (2.34, 2.07)	C^{γ}, 27.25 (1.69, 1.46); C^{δ}, * (3.65, 3.63)
V107	114.37 (8.43)	175.21	58.65 (4.60)	36.35 (1.99)	$C^{\gamma1}$, 21.06 (0.87); $C^{\gamma2}$, 20.22 (0.53)
D108	123.52 (8.44)		52.78 (4.34)	41.41 (3.02, 2.88)	
P109		177.09	65.89 (4.16)	32.00 (2.47, 2.21)	C^{γ}, 27.70 (2.03, 2.03); C^{δ}, * (3.86, 3.64)
E110	117.74 (9.23)	179.19	59.67 (4.11)	29.00 (2.03, 1.96)	C^{γ}, 36.88 (2.38, 2.38)
E111	118.70 (8.02)	179.56	57.79 (4.35)	28.87 (2.03, 2.03)	C^{γ}, 35.08 (2.28, 2.28)
I112	119.19 (7.14)	178.41	65.87 (3.52)	37.01 (1.11)	$C^{\gamma1}$, 27.19 (1.63, 1.58); $C^{\gamma2}$, 17.18 (0.88); $C^{\delta1}$, * (0.71)
E113	117.74 (7.61)	179.00	59.84 (3.95)	29.09 (2.25, 2.25)	C^{γ}, 36.25 (2.44, 2.44)
R114	117.74 (7.50)	179.35	59.43 (4.06)	30.38 (2.07, 2.07)	C^{γ}, 27.78 (1.82, 1.82); C^{δ}, 44.98 (3.32, 3.32)
I115	121.11 (7.86)	177.83	65.32 (4.07)	38.06 (1.95)	$C^{\gamma1}$, * (1.27, 1.27); $C^{\gamma2}$, 16.66 (0.95); $C^{\delta1}$, * (0.72)
L116	117.26 (8.18)	177.84	57.51 (3.96)	40.90 (1.99, 1.87)	C^{γ}, 25.82 (1.46); $C^{\delta1}$, 22.78 (0.83); $C^{\delta2}$, * (0.76)

E117	116.78 (7.65)	178.30	58.71 (4.21)	29.97 (2.58, 2.38)	C^{γ}, 36.50 (2.83, 2.83)
V118	111.47 (7.44)	176.17	61.23 (4.56)	32.14 (2.00)	$C^{\gamma 1}$, 21.15 (1.07); $C^{\gamma 2}$, 19.03 (0.84)
A119	124.01 (7.54)	176.89	52.06 (4.60)	20.19 (1.40)	
E120	121.11 (8.88)		53.97 (4.76)	30.40 (2.73, 2.73)	C^{γ}, * (3.06, 3.06)
P121			(4.52)	(2.15, 2.15)	C^{γ}, * (1.90, 1.90); C^{δ}, * (3.77, 3.34)

Figure 2A1a Derived values of ^{15}N relaxation times (violet) of MTH0776. The uncertainties in each residue are plotted as standard errors (red)

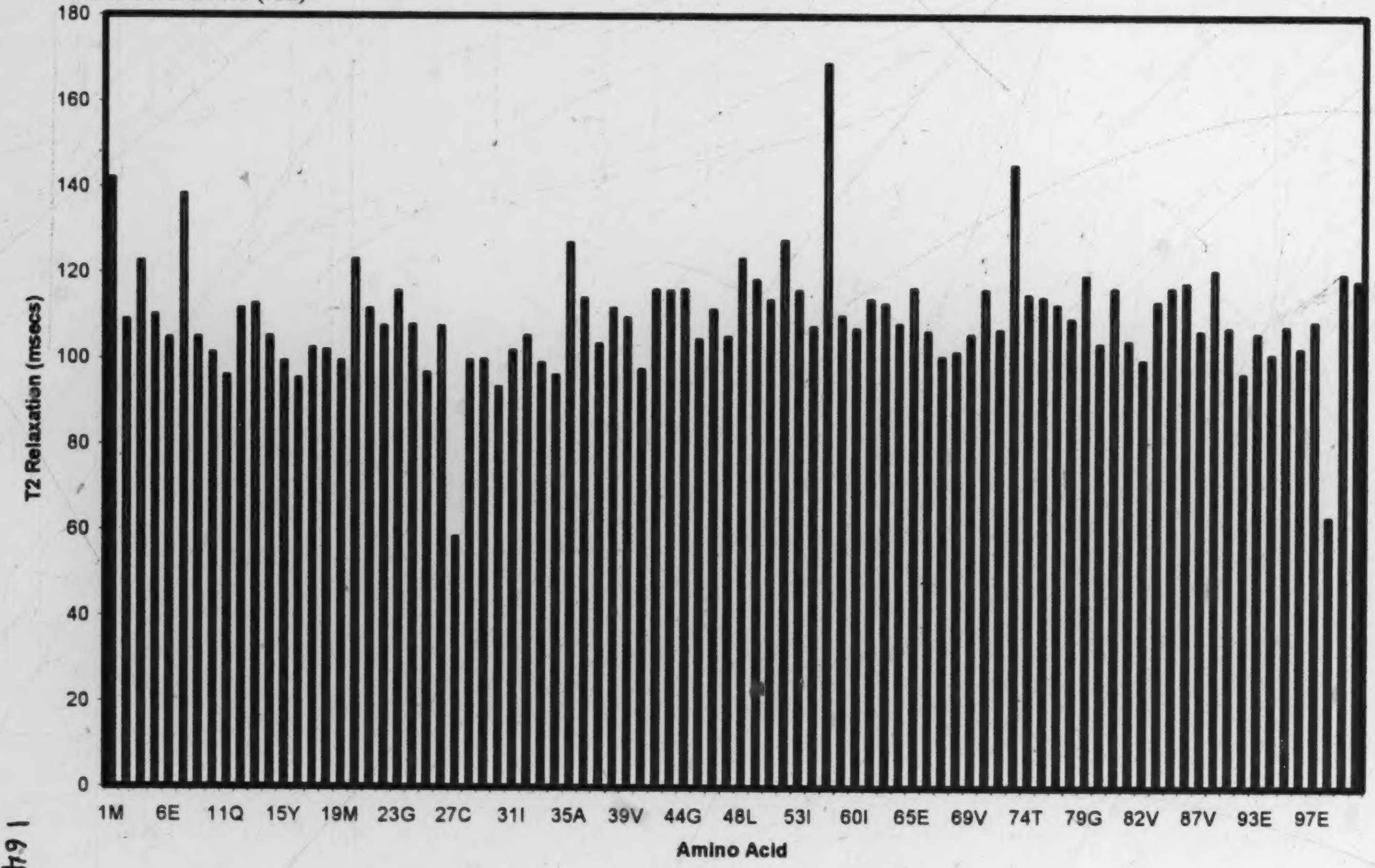


Figure 2A1b Derived values of ^{15}N relaxation times (violet) of MTH0776/MTH0777 complex. The uncertainties in each residue are plotted as standard errors (red). Note the reduced magnitude of the T_2 values of most of the residues as well as the absence of most of the residues as compared to figure A1a

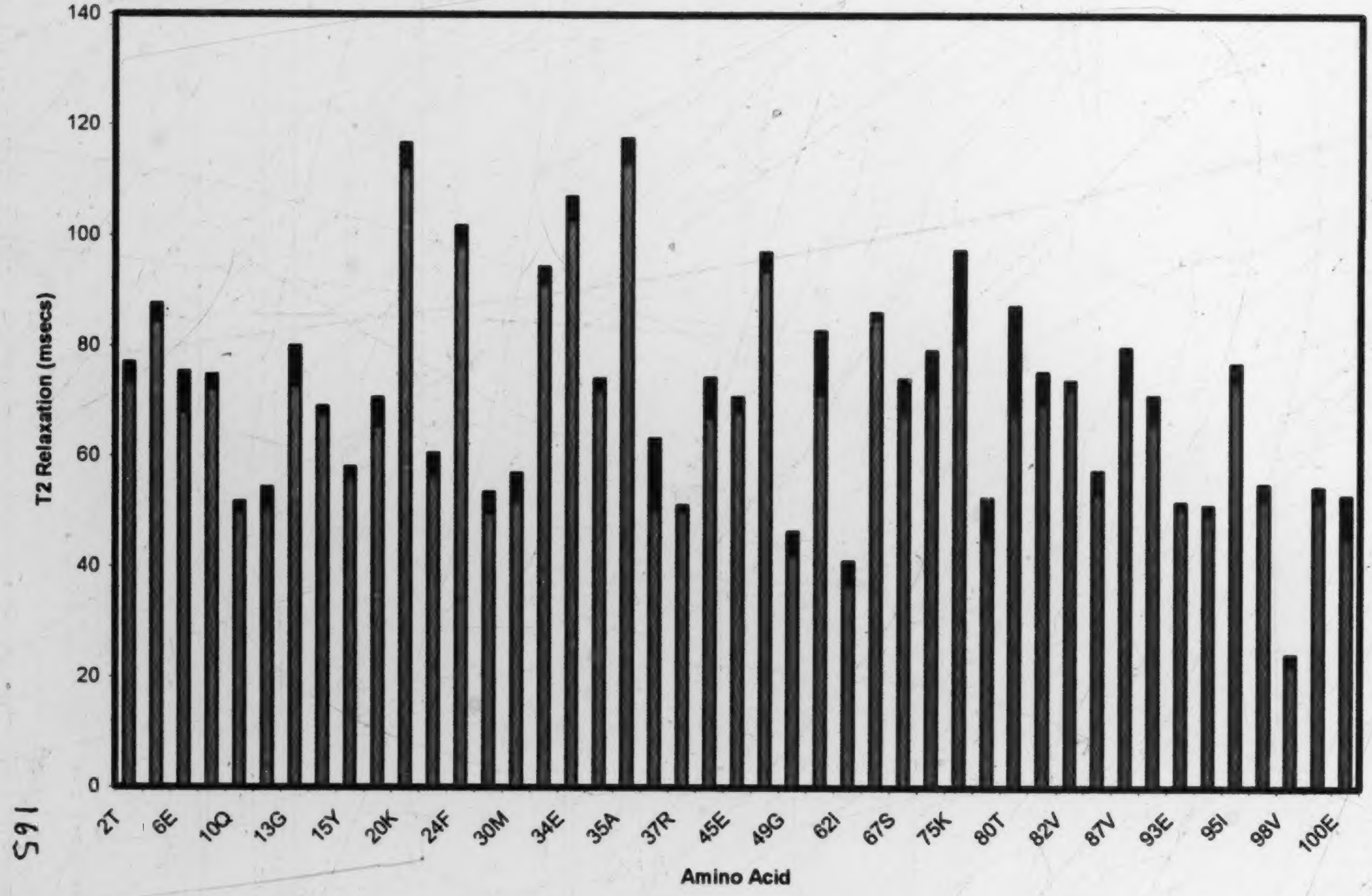


Figure 2A2 Summary of observed sequential and short –range NOEs and chemical shift index versus the sequence number. The thickness of the sequential NOEs is a qualitative indication of their intensity (thicker lines indicate stronger intensity)

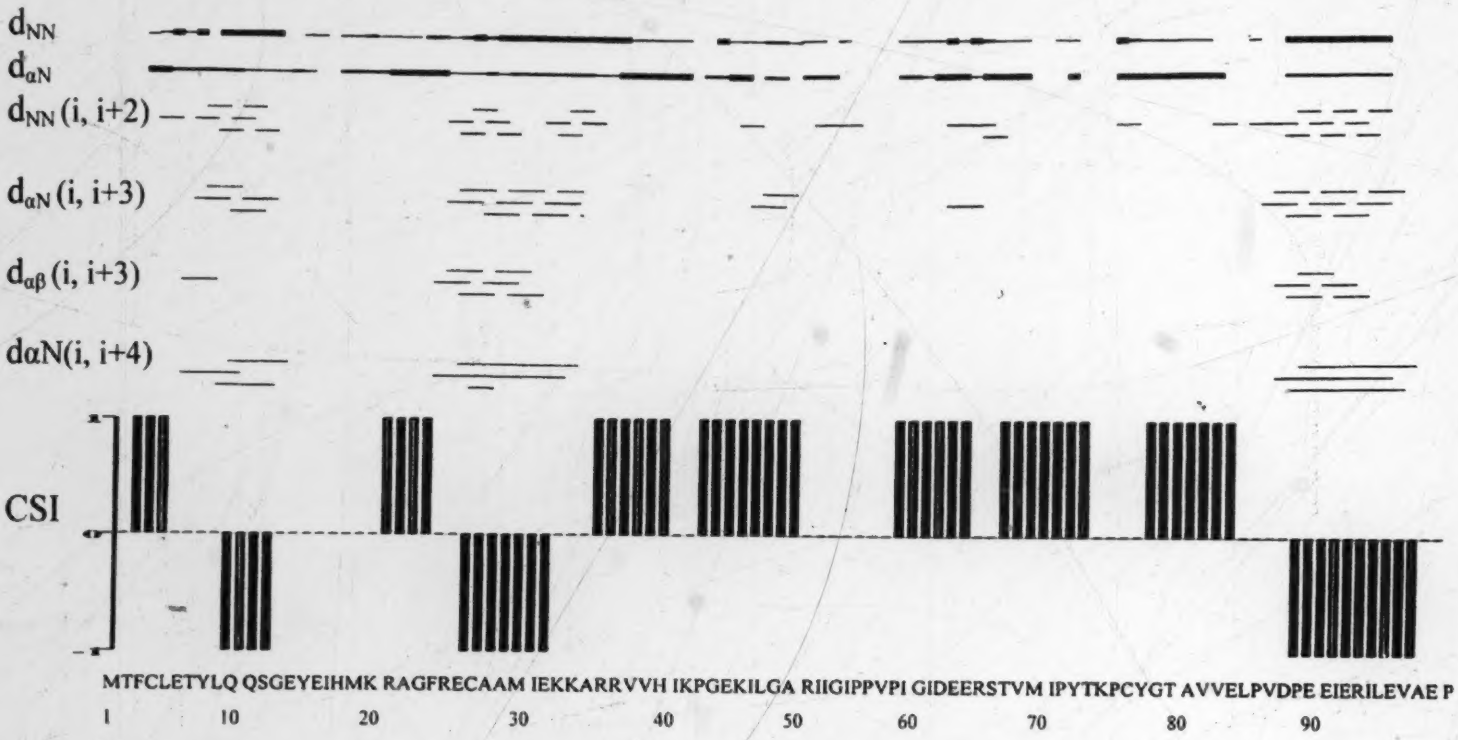
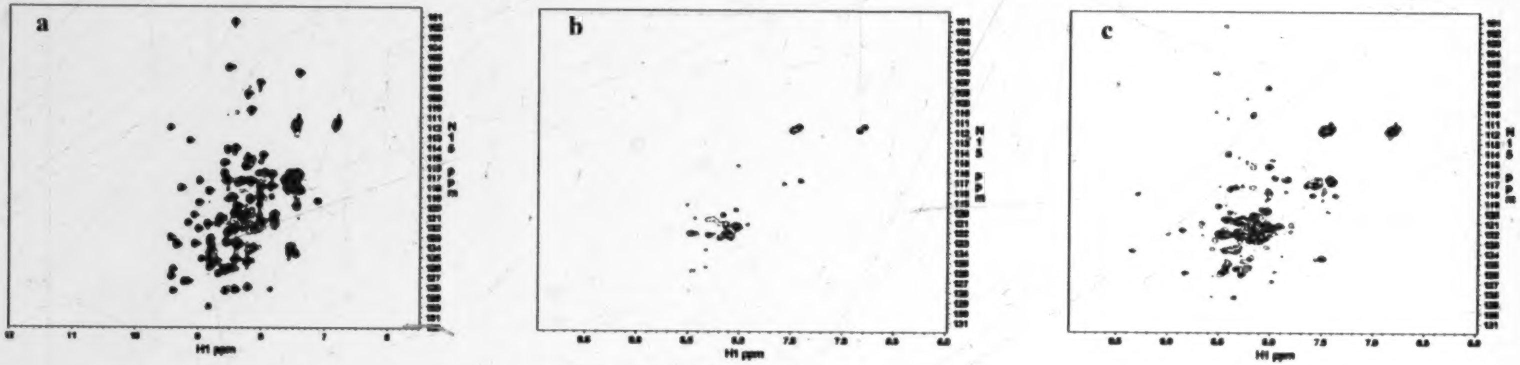


Figure 2A3 Comparison of the ^{1}H-^{15}N HSQC spectra (at different delay times) of 0.5 mM MTH0776 (a) and from 0.5 mM MTH0776 and 1 mM MTH0777 mixture (b), in 700 μL of a buffer solution pH 6.8, made up of 50 mM NaH_2PO_4, 300 mM NaCl, 50 μL D_2O, 1 mM DSS and 15 mM DTT. Spectra series (a) and (b) are rendered at the same threshold while spectra series (c) is the same as spectra series (b) but rendered at half the threshold as (b)

Delay =10miliseconds



Delay =30miliseconds

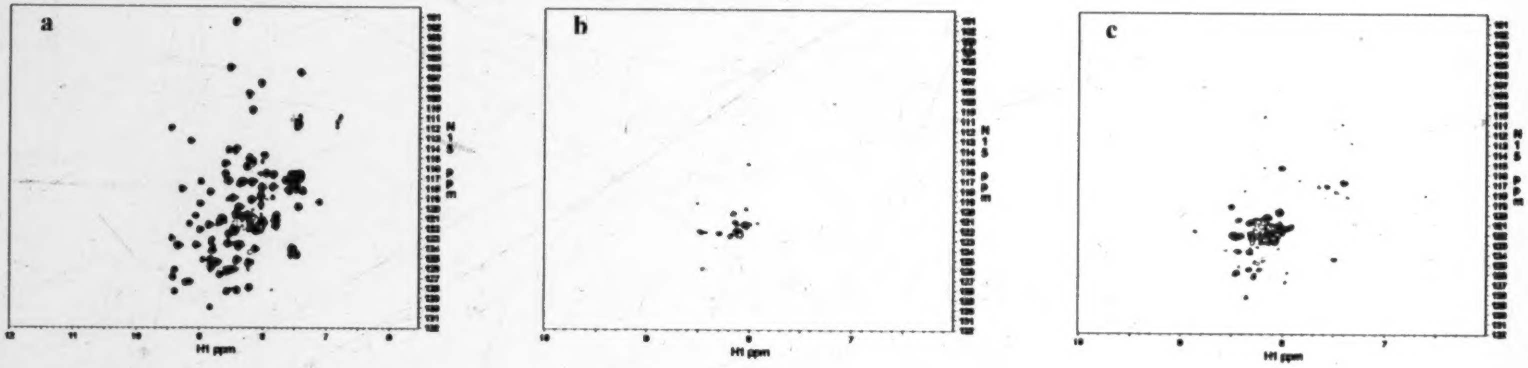
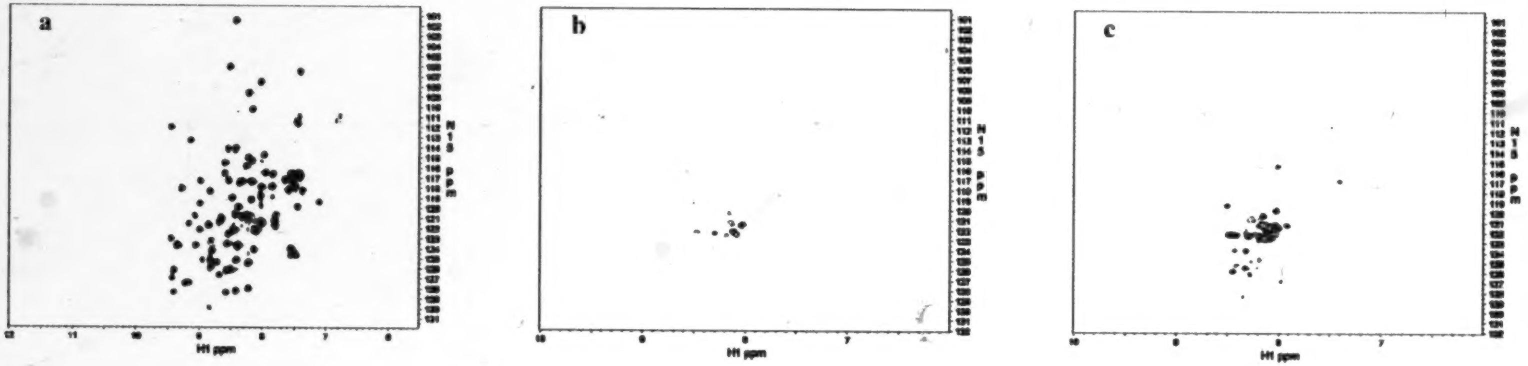


Figure 2A3 contd.

Delay =50miliseconds



Delay =70miliseconds

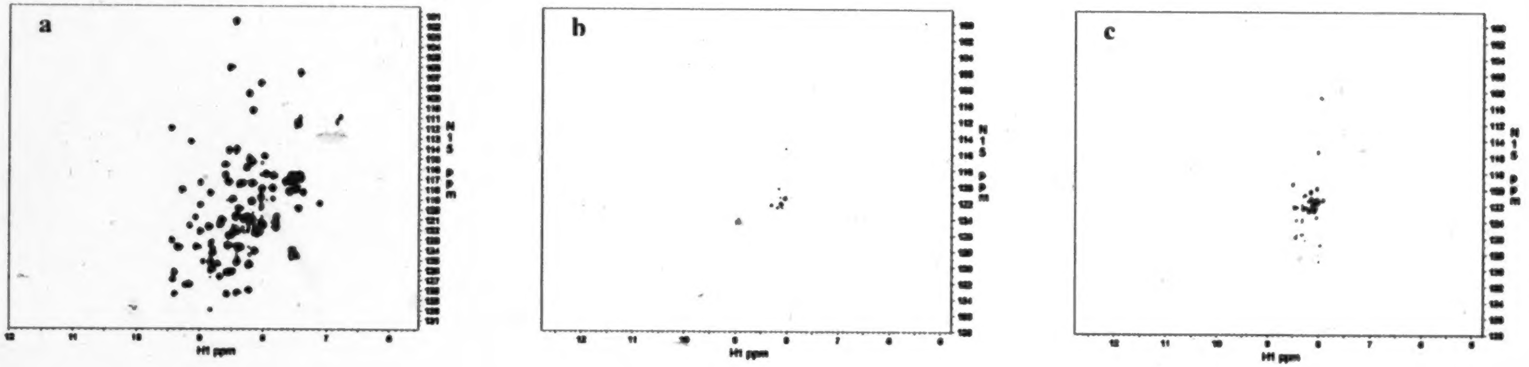
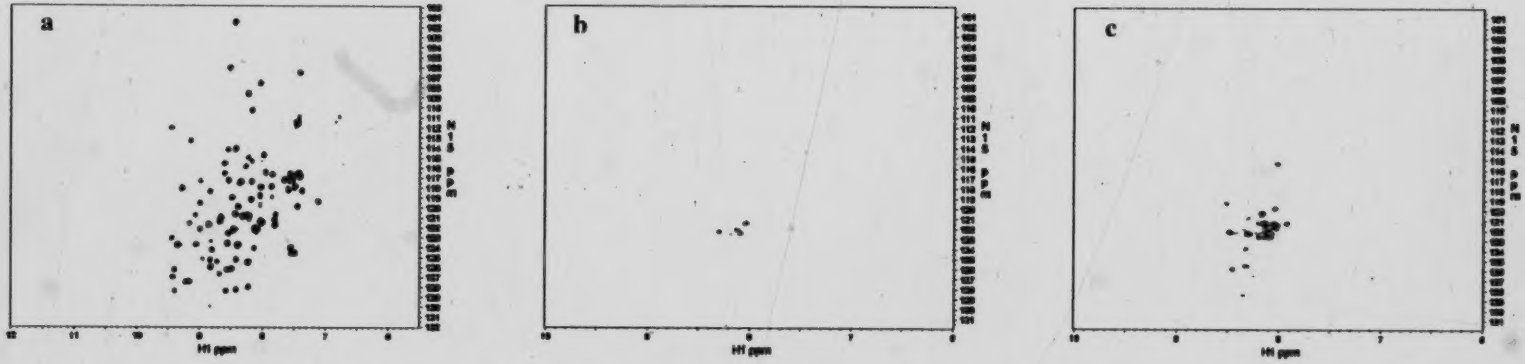


Figure 2A3 contd

Delay =90miliseconds



Delay =110miliseconds